June 1951

BIOMETRICS

Vol. 7 No. 2

JOURNAL OF THE BIOMETRIC SOCIETY

A Re-Examination of Rahn's Data on the Number of Genes in Bacteria — *D. J. Finney & L. Martin*

Rectangular Lattices and Partially Balanced Incomplete Block Designs — *K. R. Nair*

Notes on the Estimation of Gross and Net Reproduction Rates by Methods of Statistical Sampling — *J. C. Koop*

The Use of Systematic 5 x 5 Squares — *Øivind Nissen*

The Effect of Sulfamerazine on the Erythrocyte and Hemoglobin Content of Trout Blood — *James S. Gutsell*

Sur une Application de la Statistique Mathématique à la Biologie — *Maurice Fréchet*

Relationships among Characters Observed in Urea Clearance Tests — *H. Fairfield Smith*

On the Errors of Biological Assays with Graded Responses, and their Graphical Derivation — *Norman T. Gridgeman*

The Biometric Society

BIOMETRICS

FOUNDED BY THE BIOMETRICS SECTION OF THE AMERICAN STATISTICAL ASSOCIATION

TABLE OF CONTENTS

A Re-Examination of Rahn's Data on the Number of Genes in Bacteria . . . . . D. J. FINNEY & L. MARTIN 133

Rectangular Lattices and Partially Balanced Incomplete Block Designs . . . . . K. R. NAIR 145

Notes on the Estimation of Gross and Net Reproduction Rates by Methods of Statistical Sampling . . . . . J. C. KOOP 155

The Use of Systematic 5 × 5 Squares . . . . ØIVIND NISSEN 167

The Effect of Sulfamerazine on the Erythrocyte and Hemoglobin Content of Trout Blood . . . . . JAMES S. GUTSELL 171

Sur une Application de la Statistique Mathématique à la Biologie MAURICE FRÉCHET 180

Relationships among Characters Observed in Urea Clearance Tests . . . . . H. FAIRFIELD SMITH 185

On the Errors of Biological Assays with Graded Responses, and their Graphical Derivation . . . . . NORMAN T. GRIDGEMAN 200

Queries 222

The Biometric Society 224

Number 2 June 1951 Volume 7

Material for *Biometrics* should be addressed to Miss Gertrude Cox, Institute of Statistics, Box 5457, Raleigh, North Carolina, except that authors residing in one of the following organized regions can expedite the handling of their papers by submitting them to the Assistant Editor for that region.

British Region: Dr. D. J. Finney, 91 Banbury Road, Oxford, England; **Australasian Region:** Dr. E. A. Cornish, University of Adelaide, Adelaide, Australia; **French Region:** Dr. Georges Teissier, Faculte des Sciences de Paris, 1 rue V. Cousin, Paris, France

Material for Queries should go to Professor G. W. Snedecor, Statistical Laboratory, Iowa State College, Ames, Iowa.

Articles to be considered for publication should be submitted in triplicate.

THE BIOMETRIC SOCIETY

General Officers

President, Arthur Linder; *Secretary-Treasurer*, C. I. Bliss; *Council*, Maurice H. Belz, Joseph Berkson, William G. Cochran, Georges Darmois, David J. Finney, R. A. Fisher, John W. Hopkins, N. K. Jerbe, P. C. Mahalanobis, Donald Mainland, Kenneth Mather, Margaret Merrell, V. G. Panse, O. E. Sette, P. V. Sukhatme, O. Tedin, E. B. Wilson, Frank Yates.

Regional Officers

Eastern North American Region: *Vice President*, H. W. Norton; *Secretary-Treasurer*, Walter T. Federer. British Region: *Vice President*, R. A. Fisher; *Secretary*, D. J. Finney; *Treasurer*, A. R. G. Owen. Western North American Region: *Vice President*, G. A. Baker; *Secretary-Treasurer*, W. C. Rollins. Australasian Region: *Vice President*, C. W. Emmens; *Secretary-Treasurer*, J. A. Keats. Indian Region: *Vice President*, P. C. Mahalanobis; *Secretary*, C. Radhakrishna Rao; *Treasurer*, Anakul Chandra Das. French Region: *Vice President*, Maurice Fréchet; *Secretary-Treasurer*, Daniel Schwartz.

Editorial Board

Biometrics

Editor: Gertrude M. Cox, *Assistant Editors and Committee Members*, C. I. Bliss, E. A. Cornish, W. J. Dixon, John W. Fertig, D. J. Finney, O. Kempthorne, Horace W. Norton, A. M. Mood, H. Fairfield Smith, G. W. Snedecor, Georges Teissier and Jane Worcester. *Managing Editor:* Sarah P. Carroll.

The Biometric Society is an international society devoted to the mathematical and statistical aspects of biology and welcomes to membership biologists, mathematicians, statisticians and others who are interested in its objectives. Through its six regional organizations the Society sponsors regional and local meetings. National secretaries serve the interests of members in Italy, Denmark and the Netherlands and there are many members "at large". Dues in the Society for 1951 are as follows: Full membership including subscription to *Biometrics* is $7.00. Members of the Biometrics Section of the American Statistical Association who subscribe to the journal through that organization may become members of The Biometric Society on the payment of $3.00 annual dues.

Annual subscription rates to non-members are as follows: For American Statistical Association Members, $4.00; for subscribers, non-members of either American Statistical Association or The Biometric Society, $7.00. Subscriptions and applications should be sent to The Biometric Society, 52 Hillhouse Ave., New Haven, Connecticut, U.S.A.

Entered as second-class matter at the Post Office at New Haven, Conn., under the Act of March 3, 1879. Additional entry at Richmond, Va. Business Office, 52 Hillhouse Ave., New Haven, Conn. *Biometrics* is published quarterly—in March, June, September and December.

A RE-EXAMINATION OF RAHN'S DATA ON THE NUMBER OF GENES IN BACTERIA

D. J. Finney

University of Oxford

L. Martin

Institut Agronomique de Gembloux et Université de Bruxelles

1. INTRODUCTION

According to current cytogenetic theory, each chromosome in every cell of a living organism comprises a number of genes. *Inter alia*, these genes have the following properties:—

(a) They are the carriers and transmitters of hereditary characters;

(b) They control the specific enzymatic activity of the cell;

(c) They play an important part in the initiation of cell division.

This paper is concerned only with the third of these properties, and with the light it can throw on the numbers of genes in a cell. The natural process of cell division must be preceded by duplication of the constituent chromosomes, which in turn requires the duplication of every gene and is the culmination of an elaborate synthetic process. Luria (1950) has pointed out that 'There is something peculiar to homologous replication that sets it aside from other types of synthetic reactions. The replication of specific biological units must involve the building of complex specific molecules or molecular aggregates, the only permissible limitation to identity of model and replica being the production of "mutated" structures—the production, that is, of modified elements replicated in the modified form'. Claude (1949) has emphasized the difficulties in this 'template theory, elaborated to account for the supposedly unique process of gene reproduction', which 'assumes that the gene or the chromosome serves as a mold for the systematic apposition of a new substance'. He goes on to state that 'in the light of biochemical processes already known, it is conceivable that the duplication of essential and characteristic cell substances is the end

result of a series of rigidly ordered chains of reactions, the final product in turn taking part at some point, and thereby directing the specificity, of the same or other, interlocked, biochemical cycles.'

In contrast to Rashevsky's (1948) explanation of mitosis with the aid of a physico-chemical model of cytoplasmic behaviour, Rahn (1931) formulated a statistical model of the division of microorganisms in terms of gene duplication. Though his model was over-simplified, especially in that it related only to the nucleus and ignored cytoplasmic activity, it is important as indicating the type of phenomena to be expected if gene duplication is the chief controlling factor, and it ought then to be of value at least as an approximation to a more complex truth. Rahn's discussion used the following assumptions about the nature of the genes in unicellular organisms:—

(i) Each cell contains the same number, g, of genes;

(ii) Each gene has a chemical structure that distinguishes it from other genes in the same cell and that it shares with genes at the corresponding chromosomal loci in other cells (genes corresponding in this manner will be referred to as genes of the same *type*);

(iii) Gene duplication is a chemical process, and is subject to chemical laws; by analogy with the law of mass action, the rate of gene duplication at any instant is supposed proportional to the number of genes of the same type that have not yet doubled;

(iv) The duplication of any gene takes place independently of that of all others in the same cell;

(v) Cell mitosis takes place as soon as all genes in the cell have doubled;

(vi) The probability that the gene at a particular locus doubles in any specified unit time interval is the same for all gene types.

Of these assumptions, (i) would be generally accepted as a characteristic of the normal cell nucleus, (ii)–(v) are the basis of Rahn's model, and (vi) is a simplifying assumption introduced (in the absence of any indications to the contrary) as an aid to the obtaining of mathematically tractable formulae.

2. RAHN'S DISTRIBUTION OF TIMES

For a specified cell, the probability that duplication at any locus has taken place within a time t may be defined to be $p(t)$. Assumption (iii) above shows that

$$p(t) = 1 - q^t, \tag{1}$$

where q is the probability that duplication does *not* take place in unit

time; by (vi), q is a constant for all loci. By (iv) and (v), the probability that mitosis occurs within a time t, $P(t)$, is the probability that the g genes of a cell have all doubled in that time:

$$P(t) = (1 - q^t)^g. \tag{2}$$

For any possible values of g and q, $P(t)$ increases steadily from zero at $t = 0$ to unity as a limit. Since $P(t)$ is the cumulative frequency function for mitosis, the probability of mitosis in time dt is $f(t)\,dt$, where

$$f(t) = \frac{d}{dt}\{P(t)\}. \tag{3}$$

It is convenient to replace the parameter q by a new parameter α, where

$$q = e^{-\alpha}. \tag{4}$$

Hence

$$f(t) = \alpha g e^{-\alpha t}(1 - e^{-\alpha t})^{g-1}. \tag{5}$$

The probability of mitosis in any time interval is thus completely specified by two parameters, g and α. The *genetic parameter*, g, must be constant for one species of organism, whatever the conditions of experiment; the *kinetic parameter*, α, may vary from one set of observations to another.

3. METHODS OF ESTIMATING THE PARAMETERS

Equation (2) may be written

$$-\alpha t = \log (1 - P^{1/g}). \tag{6}$$

The ratio of the times required for any two specified values of P to be attained is independent of α, and could be the basis of an estimation of g in the manner proposed by Rahn. Using the agar hanging block technique, he had made counts of the times required for fission of individual bacteria in a culture of *B. aerogenes*. From this empirical assessment of the distribution $P(t)$ he read the times that elapsed before 10% and before 90% of the bacteria had divided, and estimated g from the ratio of these times with the aid of equation (6). Unfortunately, the ratio is not very sensitive to variations in g; Rahn was forced to conclude that 'hence it seems rather hopeless to determine the number of gene-types by this method if the number is high, considering also the inaccuracy in determining experimentally the order of growth'.

Ingenious though this method of estimation is, more sophisticated statistical techniques might be expected to yield better estimates of the

two parameters and also to allow tests of the adequacy of equation (2). In particular, the method of maximum likelihood is known to give estimates of maximal precision for large samples, under conditions that need not be specified here (Fisher, 1922; Cramér, 1946). Even for small samples, the estimates are likely to be more precise than any based upon equation (6), because they make full use of the data instead of employing only certain percentiles. If the recorded times to fission for n cells are t_i $(i = 1, 2, \cdots, n)$, the likelihood function, L, is the product of expressions for $f(t)$ and therefore

$$\log L = n \log \alpha + n \log g - \alpha \sum_{i=1}^{n} t_i + (g - 1) \sum_{i=1}^{n} \log (1 - e^{-\alpha t_i}). \tag{7}$$

The maximum likelihood estimates, $\hat{g}$ and $\hat{\alpha}$, are the solutions of the equations

$$\left.\begin{aligned} \frac{\partial}{\partial g} (\log L) &= 0, \\ \frac{\partial}{\partial \alpha} (\log L) &= 0. \end{aligned}\right\} \tag{8}$$

An iterative process for solving equations (8) may be set up (as for example by Finney, 1949; 1951, Appendix II), but this seems to converge rather slowly. An alternative is to write the equations as

$$\left.\begin{aligned} g &= -\frac{n}{\sum \log u_i} \\ \text{and} \\ g &= 1 + \frac{\sum t_i - \frac{n}{\alpha}}{\sum \left\{ \frac{t_i(1 - u_i)}{u_i} \right\}}, \end{aligned}\right\} \tag{9}$$

where

$$u_i = 1 - e^{-\alpha t_i}. \tag{10}$$

The right-hand sides of equations (9) may be evaluated for two trial values of α, and a new trial value obtained by interpolation so as to make the difference between these zero. Further interpolations are made until both right-hand sides agree in giving the same value, which

is $\hat{g}$. If the second differential coefficients of $\log L$ with respect to g and α are evaluated for the particular values $\hat{g}$ and $\hat{\alpha}$, and are written as a matrix

$$\begin{pmatrix} \dfrac{\partial^2 \log L}{\partial \hat{g}^2} & \dfrac{\partial^2 \log L}{\partial \hat{g}\, \partial \hat{\alpha}} \\ \dfrac{\partial^2 \log L}{\partial \hat{g}\, \partial \hat{\alpha}} & \dfrac{\partial^2 \log L}{\partial \hat{\alpha}^2} \end{pmatrix},$$

the elements of the inverse of this matrix, with signs changed, are, for large n, asymptotically equal to the variances and covariance of $\hat{g}$ and $\hat{\alpha}$ (Fisher, 1922). Whether these quantities are reasonable approximations to the variances and covariance when n is small is unknown, but the corresponding assumption is often made in other applications of maximum likelihood estimation and no alternative method of obtaining any idea of the variances is at present known. On this supposition

$$\left.\begin{aligned} V(\hat{g}) &= \left[\frac{n}{\hat{\alpha}^2} + (\hat{g} - 1) \sum \left\{\frac{t_i^2(1 - u_i)}{u_i^2}\right\}\right] \div D, \\ V(\hat{\alpha}) &= \frac{n}{\hat{g}^2} \div D, \\ \text{and} \\ C(\hat{g}, \hat{\alpha}) &= \left[\sum \left\{\frac{t_i(1 - u_i)}{u_i}\right\}\right] \div D, \end{aligned}\right\} \tag{11}$$

where

$$D = \frac{n}{\hat{g}^2}\left[\frac{n}{\hat{\alpha}^2} + (\hat{g} - 1) \sum \left\{\frac{t_i^2(1 - u_i)}{u_i^2}\right\}\right] - \left[\sum \left\{\frac{t_i(1 - u_i)}{u_i}\right\}\right]^2. \tag{12}$$

4. ANALYSIS OF DATA

Rahn (1931) reported the results of nine sets of observations on *B. aerogenes*, these being recorded as numbers of fissions occurring in successive five-minute time intervals. Professor Rahn has very kindly made available to us two additional sets of observations, those of February 16 and 23, which formed part of the same series but were omitted from his paper because they were small and showed slower rates of fission; there appears to be no reason for excluding them from the comprehensive analysis described here. The complete data are shown in Table 1. Estimates of g and α, obtained from each series by applica-

TABLE 1

NUMBERS OF FISSIONS OF *B. aerogenes* OBSERVED IN SUCCESSIVE FIVE MINUTE TIME INTERVALS

Experiment	Time interval														Total
	2	3	4	5	6	7	8	9	10	11	12	13	14	>14	
Feb. 16			2	0	4	4	5	3	5	3	1				27
Feb. 17					2	13	12	7	7	2	1				44
Feb. 23					2	6	10	12	11	5	1	0	1		48
Feb. 24			2	3	4	20	16	11	2	1	0	1			60
March 2	1	0	3	19	22	21	10	5	2	1					84
March 3		1	3	12	20	38	19	18	8	4	1	1	0	1	126
March 6			4	12	30	19	15	1	0	2	1				84
March 10			6	10	16	26	18	9	3	1	0	0	0	4	93
March 17		6	23	43	30	9	0	1							112
Nov. 12			13	60	24	3									100
Nov. 14			5	13	4	5	2	1							30
Total	1	7	61	172	158	164	107	68	38	19	5	2	1	5	808

tion of equations (9), are summarized in Table 2, the standard errors quoted being the square roots of the variances given by equations (11). In the calculations, fissions recorded as having occurred in time intervals nos. 2, 3, 4, $\cdots$ were assumed to have occurred after times $7\frac{1}{2}$,

TABLE 2

SUMMARY OF MAXIMUM LIKELIHOOD ESTIMATES OF PARAMETERS

Experiment	No. of cells, n	No. of genes, ϕ	Kinetic parameter, $\hat{\alpha}$	Rate of gene duplication per 5 minutes $(1 - e^{-\hat{\alpha}})$	Correlation coefficient between ϕ and $\hat{\alpha}$
Feb. 16	27	22.4 ± 10.3	0.473 ± 0.071	0.377	0.909
Feb. 17	44	534 ± 383	0.877 ± 0.105	0.584	0.978
Feb. 23	48	244 ± 139	0.708 ± 0.077	0.507	0.967
Feb. 24	60	63.0 ± 23.4	0.652 ± 0.061	0.479	0.937
March 2	84	29.4 ± 7.7	0.652 ± 0.052	0.479	0.910
March 3	126	37.5 ± 9.0	0.600 ± 0.041	0.451	0.929
March 6	84	89.9 ± 32.6	0.833 ± 0.070	0.565	0.953
March 10	93	31.1 ± 8.8	0.593 ± 0.050	0.447	0.931
March 17	112	59.1 ± 16.5	0.991 ± 0.071	0.629	0.941
Nov. 12	100	1,010 ± 493	1.594 ± 0.118	0.797	0.979
Nov. 14	30	91.5 ± 60.1	0.997 ± 0.151	0.632	0.961

$12\frac{1}{2}$, $17\frac{1}{2}$, $\cdots$ minutes; the grouping appeared to be sufficiently fine to make more elaborate treatment unnecessary. The five times beyond the fourteenth interval were assumed to be $72\frac{1}{2}$ minutes.

For each experiment, expected frequencies in each time interval may be calculated by inserting $\hat{g}$ and $\hat{\alpha}$ in equation (2). Table 3 summarizes these calculations for the experiment of March 3. The data conform

TABLE 3

OBSERVED AND EXPECTED FREQUENCIES FOR THE EXPERIMENT OF MARCH 3 ($\hat{g} = 37.45$, $\hat{\alpha} = 0.6001$)

Time interval (j)	Proportion doubled at end of interval [$P(5j)$]	Number of fissions in interval	
		Observed	Expected
3	0.0012	1	0.2
4	0.0284	3 } 4	3.4 } 3.6
5	0.1478	12	15.0
6	0.3546	20	26.1
7	0.5681	38	26.9
8	0.7340	19	20.9
9	0.8442	18	13.9
10	0.9113	8	8.4
11	0.9503	4	4.9
12	0.9724	1	2.8
13	0.9848	1 } 3	1.6 } 6.3
14	0.9916	0	0.9
>14	1.0000	1	1.0
Totals	—	126	126.0

well to expectation, the value of χ^2 being 9.9 with 6 degrees of freedom (nine classes were used, expected and observed totals were made to agree, and two parameters were estimated). The other series would show similar good agreement if they were tested in this way.

5. DISCUSSION OF DATA

Chief interest attaches to comparisons between the eleven experiments. As emphasized above, the true value of the genetic parameter must be the same in all experiments. At first sight, the finding of estimates of g ranging from 22 to 1,010 seems to contradict this so strongly that, in spite of the large standard errors, the hypothesis and assumptions on which equation (2) rests must be wrong. Surely any test of significance for differences in the g column of Table 2 would show

them to be far too great for explanation in terms of sampling variation! The answer, however, is not quite as simple as at first appears.

In the distribution of times represented by equation (2) or equation (5), the roles of g and α are biologically distinct but mathematically similar: each has an important influence on the location of the distribution. Evaluation of the expected mean time gives

$$E(t) = \int_0^\infty tg(t)\,dt$$
$$= \frac{1}{\alpha}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{g}\right), \tag{13}$$

which, for large g, may be written

$$E(t) \sim (\gamma + \log_e g)/\alpha, \tag{14}$$

where γ is Euler's constant, $0.5772 \cdots$; even for $g = 10$, the error in replacing (13) by (14) is less than 2%, and it rapidly decreases for larger g, so that for most practical purposes (14) may be used. Consequently, pairs of estimates of g, α which keep the right-hand side of (14) constant will be equally satisfactory as judged by the first moment of the distribution; only the dispersion in the times can discriminate between effects of changes in g and of changes in α. This property of the distribution is evident during the process of maximum likelihood estimation: a considerable change in the value of one parameter can be made without seriously departing from the conditions expressed by equations (9), provided only that it is compensated by a change of the other parameter in the same direction. In fact, the covariance of the estimates is large relative to the variances; a correlation coefficient may be defined as

$$r_{\hat{g}.\hat{\alpha}} = \frac{C(\hat{g}, \hat{\alpha})}{\{V(\hat{g})\,V(\hat{\alpha})\}^{\frac{1}{2}}}, \tag{15}$$

and is found to exceed 0.9 for every experiment (Table 2). Moreover, the eleven pairs of estimates in Table 3 show the same tendency—a strong positive correlation.

At our invitation, Mr. R. M. Jones has made a further study of the properties of the distribution represented by equation (5). He has shown that the cumulants have a simple asymptotic form when g is large:

$$\kappa_i \sim \frac{(i-1)!}{\alpha^i} \sum_{s=1}^{\infty} s^{-i} \qquad (i \geq 2), \tag{16}$$

κ_1 being given by (14). Even when g is only of the order of 30, these expressions are very near to the truth, and thus all cumulants except the first are nearly independent of g. In particular, the variance of the distribution of times is

$$V(t) \sim \pi^2/6\alpha^2. \tag{17}$$

The first two moments of a sample of n observations might therefore be used for giving α_0 , g_0 , estimates of the two parameters according to the 'method of moments'. These would be defined by

$$\left.\begin{aligned} \alpha_0 &= \pi\left\{\frac{n-1}{6\sum_1^n (t-\bar{t})^2}\right\}^{\frac{1}{2}}, \\ g_0 &= \frac{\alpha_0 \sum_1^n (t)}{n} - \gamma. \end{aligned}\right\} \tag{18}$$

Calculation of α_0 and g_0 is easier than that of $\hat{\alpha}$ and $\hat{g}$. In nine of the experiments, α_0 is greater than $\hat{\alpha}$, and, as is to be expected, g_0 then exceeds $\hat{g}$. The values of α_0 range from 0.605 to 1.879, and those of g_0 from 28.2 to 3,630. The method of moments appears to be appreciably less efficient than the method of maximum likelihood, however, since the sampling variance of α_0 may be shown to be approximately

$$V(\alpha_0) \simeq \frac{1.1\alpha^2}{n}, \tag{19}$$

and the square root of this always exceeds the standard error of $\hat{\alpha}$ in Table 2.

In view of the correlation between g and $\hat{\alpha}$, it is natural to inquire whether, despite their seeming discrepancies, the experiments may not agree satisfactorily with predictions based upon a common value of g. The value of α need not be the same for all, and the ideal estimation procedure would estimate g and eleven separate values of α from all the data. This would be exceedingly laborious, and, as an alternative, the more restricted theory that α also must be constant has been tried. Estimates of g and α found by applying equations (9) to the total line of Table 1 are

$$\hat{g} = 25.95 \pm 2.39$$

$$\hat{\alpha} = 0.6139 \pm 0.0177,$$

with

$$r_{\hat{g},\hat{\alpha}} = 0.924.$$

These estimates are only a little greater than the smallest recorded in Table 2; the explanation lies in the fact that the small values in that Table have smaller variances than the large, and therefore exert a dominant influence when the data are pooled. As confirmation that the smallness of $\hat{g}$ was not a consequence of serious bias in the maximum likelihood process, estimates by the method of moments were also formed, and are

$$g_0 = 34.2,$$

$$\alpha_0 = 0.657 \pm 0.024.$$

These are appreciably larger than $\hat{g}$ and $\hat{\alpha}$, but of the same order of magnitude; judging by α_0, they are less precise estimates. Table 4 is analogous to Table 3 and shows the observed and expected frequencies

TABLE 4

OBSERVED AND EXPECTED FREQUENCIES FOR THE POOLED DATA
($\hat{g} = 25.95$, $\hat{\alpha} = 0.6139$)

Time interval (j)	Proportion doubled at end of interval [$P(5j)$]	Number of fissions in interval	
		Observed	Expected
2	0.0001	1|8	0.1|9.1
3	0.0113	7|	9.0|
4	0.0975	61	69.6
5	0.2911	172	156.4
6	0.5165	158	182.1
7	0.7008	164	148.9
8	0.8255	107	100.8
9	0.9016	68	61.5
10	0.9455	38	35.5
11	0.9701	19	19.9
12	0.9837	5	11.0
13	0.9912	2	6.1
14	0.9952	1|6	3.2|7.1
>14	1.0000	5|	3.9|
Totals	—	808	808.0

for the pooled data; the value of χ^2 is 14.9 with 9 degrees of freedom. This scarcely exceeds the 10% significance level (14.7), and does not

suggest any serious disagreement between theory and observation. If the unnecessary restriction on α were dropped, and each experiment were allowed its own kinetic parameter, the agreement would presumably be improved; the improvement might not be very great, because of the correlation between estimates.

6. SUMMARY AND CONCLUSIONS

Rahn suggested a theory of cell multiplication, in terms of the duplication of all genes in a cell, according to which the frequency of mitosis depends upon the number of genes per cell and a parameter representing speed of gene duplication (for simplicity, supposed constant for all genes). The distribution of times to fission that results from this theory has been discussed; a maximum likelihood procedure for the estimation of the parameters from observational data has been established.

On the assumption that Rahn's data for *B. aerogenes* are random samples of the times that elapse before fission of individual bacteria, estimates of the parameters have been formed from his eleven experiments. All agree satisfactorily with the theory; though the separate experiments show large differences in their estimates, when all the data are combined they agree well with the theory if the number of genes per cell is about 26 and the rate of gene duplication is about 46% per five minutes. The theory actually tested has been unduly restrictive in demanding that the kinetic parameter be the same for all gene-types and on all occasions. Lesser requirements would obviously still be in accord with Rahn's physicochemical hypothesis, but the data are insufficient for investigation of whether q is really constant on all occasions and for all genes. Though statistical tests disclose no significant departure from the predictions based on equation (2), the estimate of the number of genes is so remarkably small as itself to cause suspicion. In correspondence with one of us, Professor Rahn commented 'I am as surprised about the small number of genes that you found as I was about my own results. To the uninitiated, a bacterium may seem a very simple organism, but to the bacteriologist it seems almost unbelievable that 30 genes could control all the many properties of such a cell. There is not only the ability to make cell wall and protoplasm, but the ability to decompose carbohydrates to various acids, carbon dioxide, and hydrogen, and above all to construct all of the catalysts needed for growth, e.g. the mechanism to make various amino acids from glucose and ammonia and to make proteins and nucleoproteins out of these amino acids. It must be that each gene controls a number of synthetic properties.' The only alternative to the view expressed by

Professor Rahn's last sentence appears to be that the theory expressed by equation (2) is seriously wrong in the importance it attaches to the number of genes as a factor determining rate of fission, but that the data analysed happen to agree with it as well as with the unknown correct theory.

The problem of estimating the number of genes in a micro-organism is essentially biometric in character (cf. Cochran, 1950; Martin, 1949), involving the formulation of a mathematical model of certain biological phenomena and the submission of this model to experimental test. Unfortunately, the mathematical model appropriate to Rahn's theory is such that a reasonably sensitive test of its adequacy is impossible without either vastly more data than he obtained or experimentation of an entirely different character. This conclusion, the clearest that emerges from the present paper, can give little satisfaction to statisticians or to experimenters, and, in the face of it, the numerical estimates must be accepted with reserve.

ACKNOWLEDGEMENTS

We wish to express our gratitude to Professor Otto Rahn for so freely placing additional data at our disposal, to Mr. R. Morley Jones for a mathematical study of the frequency distribution, and to Miss I. M. Strong and Miss H. M. Wild whose painstaking calculations were a much greater labour than might appear from Table 2.

REFERENCES

Claude, A. Proteins, lipids and nucleic acids in cell structures and functions. *Advances in Protein Chemistry*, 5, 423–440, 1949.

Cochran, W. G. The present status of biometry. *Bulletin de l'Institut International de Statistique*, 32, 132–150, 1950.

Cramér, H. *Mathematical Methods of Statistics*. Princeton: University Press, 1946.

Finney, D. J. The estimation of the parameters of tolerance distributions. *Biometrika*, 36, 239–256, 1949.

Finney, D. J. *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve*. (2nd edition) London: Cambridge University Press, 1951.

Fisher, R. A. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, A 222, 309–368, 1922.

Luria, S. E. Bacteriophage: An essay on virus reproduction. *Science*, 111, 507–511, 1950.

Martin, L. Evolution de la Biométrie, de Quetelet au 2me Congrès International de Biométrie (Genève, 1949). *Bulletin de l'Institut Agronomique et des Stations de Recherches de Gembloux, Belgique*, 17, 43–66, 1949.

Rahn, O. A chemical explanation of the variability of the growth rate. *Journal of General Physiology*, 15, 256–277, 1931.

Rashevsky, N. *Mathematical Biophysics* (2nd edition). Chicago: University of Chicago Press, 1948.

RECTANGULAR LATTICES AND PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS

K. R. NAIR

Forest Research Institute, Dehra Dun, India

HARSHBARGER (1946, 1947, 1949) developed a set of incomplete block designs in which the number of varieties (or treatments) is the product of two consecutive integers and the number of replications of each variety is either 2 or 3 and their multiples. He called them *simple* and *triple rectangular lattices* respectively.

It is the purpose of this paper to show that the simple rectangular lattice belongs to the category of partially balanced incomplete block designs developed by Bose and Nair (1939) and later modified by Nair and Rao (1942). It will also be shown that the triple rectangular lattice does not, in general, satisfy the combinatorial conditions required of a p.b.i.b. design.

A brief reference to these conditions had been made in this journal by Bose (1947). They may be stated as follows:—

(1) There are v varieties (or treatments) each replicated in r plots. The vr plots are arranged on the ground in b blocks each containing k plots ($k < v$). No treatment will occur more than once in a block.

(2) With respect to any treatment θ the remaining ones can be divided into m groups of $n_1, n_2, \cdots, n_m$ each, so that each treatment of the i-th group appears with θ in λ_i blocks. The treatments of the i-th group are said to be i-th associates of θ.

(3) The relation of association is symmetrical. Thus if the treatment θ is an i-th associate of ϕ, then ϕ is an i-th associate of θ. Further, if p_{jk}^{i} is the number of treatments which are both j-th associates of θ and k^*-th associates of ϕ, then the value of p_{jk}^{i} is independent of the pair of i-th associates θ and ϕ with which we start.

The following relations exist:—

*This k is not to be confused with the number of plots per block which is also denoted by k.

$$vr = bk \tag{1}$$

$$\sum_{1}^{m} n_i \lambda_i = r(k-1) \tag{2}$$

$$p_{jk}^{i} = p_{kj}^{i} \tag{3}$$

$$\sum_{k=1}^{m} p_{jk}^{i} = n_j - 1 \quad \text{or} \quad n_j \quad \text{according as } i = j \quad \text{or} \quad i \neq j \tag{4}$$

$$n_i p_{jk}^{i} = n_j p_{ik}^{j} = n_k p_{ij}^{k} \tag{5}$$

Originally, Bose and Nair (1939) had imposed the condition that $\lambda_1, \cdots, \lambda_m$ were unequal, but Nair and Rao (1942) found that this condition was not necessary.

The treatment differences are estimated with m different accuracies. The 'balanced incomplete block design' can be treated as a special case of p.b.i.b. designs, with $m = 1$. The square and cubic lattices for p^2 and p^3 varieties respectively, in blocks of p plots, are special cases of p.b.i.b. designs, with $m = 2$ and 3. Many illustrative examples of these designs for which m is 2 or 3 have been given by Bose and Nair (1939). It was thought at the time that for practically useful designs m should not exceed 3.

The *simple rectangular lattice* for $p(p-1)$ varieties is a fine example of a p.b.i.b. design for which $m = 4$ when $p \geq 4$, and $m = 3$ when $p = 3$. Since only two replications are required for this design its practical usefulness is obvious.

Following Harshbarger's method, each of the $p(p-1)$ varieties will be identified by a couple of digits x and y ($x \neq y$) each taking the values 1, 2, $\cdots$ p. All varieties having the same value for x will form the x-th block of the first replication and all varieties having the same value for y will form the y-th block of the second replication.

For any particular variety (xy) there are $2(p-2)$ varieties appearing with it in one or the other of the two blocks in which it is replicated. Each of these $2(p-2)$ varieties has one digit in common with variety (xy). They will be called 1st associates of (xy) with $\lambda_1 = 1$. Hence $n_1 = 2(p-2)$.

The remaining $p^2 - 3p + 3$ varieties can be grouped into three (if $p \geq 4$) according as they have 0, 1 or 2 digits in common with (xy). They may be called *2nd*, *3rd*, and *4th* associates of (xy) with $\lambda_2 = 0$,

$\lambda_3 = 0$, $\lambda_4 = 0$ and $n_2 = (p-2)(p-3)$, $n_3 = 2(p-2)$ and $n_4 = 1$ respectively.

Values of p_{ik}^i can be easily worked out. Their values and those of the parameters obtained above are given below for $p \geq 4$.

$$v = p(p-1), \quad k = (p-1), \quad r = 2, \quad b = 2p$$

$$\lambda_1 = 1 \qquad \lambda_2 = 0 \qquad \lambda_3 = 0 \qquad \lambda_4 = 0$$

$$n_1 = 2(p-2), \quad n_2 = (p-2)(p-3), \quad n_3 = 2(p-2), \quad n_4 = 1.$$

$$p_{ik}^1 = \begin{pmatrix} (p-3) & (p-3) & 1 & 0 \\ (p-3) & (p-3)(p-4) & (p-3) & 0 \\ 1 & (p-3) & (p-3) & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$p_{ik}^2 = \begin{pmatrix} 2 & 2(p-4) & 2 & 0 \\ 2(p-4) & (p-4)(p-5) & 2(p-4) & 0 \\ 2 & 2(p-4) & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$p_{ik}^3 = \begin{pmatrix} 1 & (p-3) & (p-3) & 1 \\ (p-3) & (p-3)(p-4) & (p-3) & 1 \\ (p-3) & (p-3) & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$p_{ik}^4 = \begin{pmatrix} 0 & 0 & 2(p-2) & 0 \\ 0 & (p-2)(p-3) & 0 & 0 \\ 2(p-2) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

When $p = 3$, n_2 becomes zero so that m reduces to 3. The *3rd* and *4th* associates for the case $p \geq 4$ may be called *2nd* and *3rd* associates respectively for the case $p = 3$. The parameters are

$$v = 6, \quad k = 2, \quad r = 2, \quad b = 6.$$

$$\lambda_1 = 1 \quad \lambda_2 = 0 \quad \lambda_3 = 0$$

$$n_1 = 2, \quad n_2 = 2, \quad n_3 = 1$$

$$p^1_{ik} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}; \quad p^2_{ik} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}; \quad p^3_{ik} = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The statistical analysis appropriate for p.b.i.b. designs was discussed by Bose and Nair (1939) taking into account only intra-block information. Methods for using inter-block in addition to intra-block information have been given by Nair (1944) and Rao (1947).

For the simple rectangular lattice, Harshbarger directly derived expressions for the various sums of squares in the analysis of variance and for variances of treatment differences. Since this lattice has now been shown to be only a special case of p.b.i.b. designs these expressions can be obtained by substitution in the general expressions for the p.b.i.b. designs.

Triple rectangular lattice. I had naturally expected that, like the simple rectangular lattice, the triple rectangular lattice discussed by Harshbarger (1949) will also be a special case of p.b.i.b. designs, with m taking the value 7 in general. But this expectation has not come true except in the cases $v = 2 \times 3$ and $v = 3 \times 4$ when they can be shown to be special cases of p.b.i.b. designs with m equal to 2 and 3 respectively.

In the triple rectangular lattice each of the $p(p - 1)$ varieties is identified by a set of 3 digits xyz ($x \neq y \neq z$), each taking the values 1 to p. All varieties having the same value for x will form the x-th block of the first replication, all varieties having the same value for y will form the y-th block of the second replication and all varieties having the same value for z will form the z-th block of the third replication.

For any particular variety (xyz) there are $3(p - 2)$ varieties appearing with it in one or the other of the three blocks in which it is

replicated. These $3(p-2)$ varieties can be grouped into three (if $p \geq 5$) according as they have 1, 2 or 3 digits in common with (xyz). If $p = 4$, the number of groups will be two according as the number of common digits is 2 or 3. If $p = 3$, there is only one group, as all the 3 digits are common for every variety.

The remaining $p^2 - 4p + 5$ varieties can be grouped into four (if $p \geq 6$) according as they have 0, 1, 2 or 3 digits in common with (xyz). If $p = 5$, there can be no two varieties with no digits in common and hence the number of groups reduces to three. Similarly when $p = 4$, there will be only two groups, namely those with 2 or 3 digits in common. Lastly, when $p = 3$ there is only one group as all varieties have all the 3 digits alike.

In general, therefore, when $p \geq 6$, with respect to any variety (xyz), the remaining varieties apparently fall in seven groups.

Let us take the smallest triple rectangular lattice, namely, that for which $p = 3$. The six varieties may be written in the form

×	123	132
213	×	231
312	321	×

It can easily be seen that this is a p.b.i.b. design with following parameters:—

$$v = 6, \qquad k = 2, \qquad r = 3, \qquad b = 9.$$

$$\lambda_1 = 1 \qquad \lambda_2 = 0$$

$$n_1 = 3 \qquad n_2 = 2$$

$$p_{ik}^1 = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \qquad p_{ik}^2 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

Let us next take the case $p = 4$. The 12 varieties may be written

in the form

×	123	134	142
214	×	231	243
312	324	×	341
413	421	432	×

We notice that with respect to any variety all the varieties which appear with it in the same block have 2 digits alike; and among the varieties which do not so appear some have 2 digits alike and others have 3 digits alike. If we call these groups the 1*st*, 2*nd* and 3*rd* associates respectively, the above triple rectangular lattice is a p.b.i.b. design with following parameters:—

$$v = 12, \qquad k = 3, \qquad r = 3, \qquad b = 12.$$

$$\lambda_1 = 1 \qquad \lambda_2 = 0 \qquad \lambda_3 = 0$$

$$n_1 = 6 \qquad n_2 = 3 \qquad n_3 = 2$$

$$p_{ik}^1 = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}; \quad p_{ik}^2 = \begin{pmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 0 \end{pmatrix}; \quad p_{ik}^3 = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Harshbarger (1949) gives a numerical illustration of this lattice. On p. 11 of his paper five different standard errors are worked out of which three are said to be for differences between the means of two varieties occurring together in the same block and

(1) having one digit alike in the subscripts;
(2) having two digits alike in the subscripts;
(3) having three digits alike in the subscripts.

It is clear from the discussion above that groups (1) and (3) are non-existent.

The remaining two standard errors worked out by Harshbarger are said to be for differences between the means of two varieties not occurring together in the same block and

(1) having one digit alike in the subscripts;
(2) having two digits alike in the subscripts.

Of these, (1) is non-existent. Furthermore among varieties not occurring together in the same block there are pairs having three digits alike which have not been taken into account by Harshbarger in the evaluation of standard errors.

We will now examine the triple rectangular lattice for $p = 5$ and show that it does not satisfy the condition for a p.b.i.b. design. The 20 varieties may be written in the form

×	124	132	145	153
214	×	235	243	251
312	325	×	341	354
415	423	431	×	452
513	521	534	542	×

Take variety 124. The remaining 19 varieties can be grouped into five classes as follows:—

(a) varieties appearing with 124 in the same block and having either 1, 2, or 3 digits alike;
(b) varieties not appearing with 124 in the same block and having either 1 or 2 digits alike.

The varieties falling in the five groups are:—

Group: a(1)	a(2)	a(3)	b(1)	b(2)
153	132	214	235	243
325	145		513	251
354	423			312
534	521			341
				415
				431
				452
				542

If this lattice was a p.b.i.b. design these five groups would have respectively formed the 1*st*, 2*nd*, 3*rd*, 4*th* and 5*th* associate classes of variety 124 and we would have had $\lambda_1 = \lambda_2 = \lambda_3 = 1$, $\lambda_4 = \lambda_5 = 0$ and $n_1 = n_2 = 4$, $n_3 = 1$, $n_4 = 2$, $n_5 = 8$.

Assuming that it is a p.b.i.b. design let us examine the values of the parameters p_{jk}^{i}. Take, for instance, variety 534 which is a 1*st* associate of 124. The varieties belonging to each associate class of 534 are:

Associate No. 1	2	3	4	5
521	513	354	251	145
132	542		312	153
124	235			243
214	431			325
				341
				415
				423
				452

We can now evaluate the values of p_{jk}^{1}. They are

$$p_{jk}^{1} = \begin{pmatrix} 0 & 0 & 1 & 0 & 2 \\ 2 & 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 4 \end{pmatrix}$$

We find that $p^1_{ik} \neq p^1_{ki}$ for some values of j and k. Hence condition (3) of a p.b.i.b. design is not fulfilled. By successively taking a variety each from the *2nd, 3rd, 4th* and *5th* associate class of 124 and trying to evaluate p^2_{ik} , p^3_{ik} , p^4_{ik} and p^5_{ik} it can be shown that condition (5) of a p.b.i.b. design also breaks down for the 4×5 triple rectangular lattice.

When we come to 5×6 lattice further asymmetries appear in the design. For instance, consider the 30 varieties numbered as shown below:

×	123	134	145	156	162
216	×	231	243	254	265
315	326	×	341	352	364
412	425	436	×	451	463
513	524	532	546	×	561
614	621	635	642	653	×

It can be seen, for instance, that the number of varieties having 3 digits alike with variety (xyz) is either 0, 1 or 2 and *not constant.* Examples of the 3 types are $(xyz) = 243$, 123 and 254 respectively.

Take another 5×6 lattice where the z-digits are arranged in a different way as is shown in the next diagram. Here the number of varieties having 3 digits alike with variety (xyz) is either 0 or1 . Examples are $(xyz) = 241$ and 126 respectively.

Such asymmetries also appear in lattices for $p = 8$ and 10 listed by Cochran and Cox (1950).

We had seen that the number of varieties having 3 digits alike with variety (xyz) is 1 when $p = 5$. This is true also for larger odd values of p. Hence the number of groups of varieties with respect to any variety (xyz) is never 7 when p is an odd integer.

×	126	132	143	154	165
213	×	235	241	256	264
314	325	×	346	351	362
415	423	436	×	452	461
516	521	534	542	×	563
612	624	631	645	653	×

The fact that the triple rectangular lattices for $p \geq 5$ are not p.b.i.b. designs seems to show that there is a more general class of partially balanced designs with less stringent combinatorial conditions than those imposed in the p.b.i.b. designs developed by Bose, Nair and Rao. The combinatorial conditions required for this more general class of designs have not been looked into.

REFERENCES

Bose, R. C. and Nair, K. R. Partially balanced incomplete block designs. *Sankhya*, 4, 337–372, 1939.

Bose, R. C. Recent work on incomplete block designs in India. *Biometrics*, 3, 176–178, 1947.

Cochran, W. G. and Cox, G. M. *Experimental Designs*, John Wiley & Sons, Inc., New York, 1950.

Harshbarger, B. Preliminary report on the rectangular lattices. *Biometrics Bulletin*, 2, 115–119, 1946.

Harshbarger, B. Rectangular lattices, *Virginia Agricultural Experiment Station, Memoir* 1, 1947.

Harshbarger, B. Triple rectangular lattices. *Biometrics*, 5, 1–13, 1949.

Nair, K. R. and Rao, C. R. A note on partially balanced incomplete block designs, *Science and Culture*, 7, 568–569, 1942.

Nair, K. R. The recovery of inter-block information in incomplete block designs. *Sankhya*, 6, 383–390, 1944.

Rao, C. R. General methods of analysis for incomplete block designs. *Journal of the American Statistical Association*. 42, 541–561, 1947.

NOTES ON THE ESTIMATION OF GROSS AND NET REPRODUCTION RATES BY METHODS OF STATISTICAL SAMPLING

J. C. Koop

Directorate of Labour, Rangoon, Burma

1. INTRODUCTION

GROSS AND NET reproduction rates are usually computed from data on births and deaths relating to the total population. But if the data are faulty the degree to which such figures are correct cannot be assessed. The purpose of these notes is to point out that these rates may also be estimated from data, possible to obtain, by methods of statistical sampling which also supply methods of estimating their accuracy. In the case of the latter index, supplementary actuarial information in the form of female survival rates, l_x , for specific ages are necessary.

If, at a given time period, in any geographical region the entire population of women is considered, the gross reproduction rate G, according to Kuczynski, is simply the sum of the specific fertility rates for female births p_x , where the subscript x denotes, the age-group $(x - \frac{1}{2}, x + \frac{1}{2})$ of the mother. The net reproduction rate R invented by Böckh (D.V.G., 1947), will be the sum of the products of the age-specific fertility rates and each corresponding female survival rate. We need only be concerned with those groups included in the reproductive period which is usually between 15 and 50 years. In symbols

$$G = \sum_{x=15.5}^{49.5} p_x \qquad \text{and} \qquad R = \sum_{x=15.5}^{49.5} l_x p_x$$

These formulae are frequently used by demographers. In this problem, each p_x may be regarded as a population parameter to be estimated by sampling, and each l_x as a known constant.

2. SAMPLING FROM A FEMALE POPULATION STRATIFIED IN AGE-GROUPS

Suppose in a population, all females between 15 and 50 years are located, listed and stratified in yearly age-groups. Such an event can be realized when a census is being taken. Let N_x be the number of females in the age-group $(x - \frac{1}{2}, x + \frac{1}{2})$. Treat each female as a samp-

ling unit and assign to each an equal probability of being selected. If random samples of n_x females are selected from each age-group, and m_x are found to be mothers of female infants under one year in the age-group $(x - \frac{1}{2}, x + \frac{1}{2})$ then an unbiassed estimate p'_x of p_x will be given by $p'_x = m_x/n_x$. Hence unbiassed estimates G' of G and R' of R will be

$$G' = \sum_x p'_x \qquad \text{and} \qquad R' = \sum_x l_x p'_x$$

Needless to say these estimates will be subject to sampling error. We now proceed to find their variances which measure these errors. With each female, we may associate a characteristic random variable which assumes the value one, if the female in question is a mother of baby girl of age less than one year, and zero if otherwise. With this notion, and also remembering that variables not in the same age-group are mutually independent, but those within dependent, it can be easily shown that the variances of G' and R' are

$$V(G') = \sum_x \frac{p_x(1 - p_x)}{n_x} \left(\frac{N_x - n_x}{N_x - 1}\right), \tag{1}$$

and

$$V(R') = \sum_x l_x^2 \frac{p_x(1 - p_x)}{n_x} \left(\frac{N_x - n_x}{N_x - 1}\right). \tag{2}$$

Their estimates are:

$$\sum_x \frac{p'_x(1 - p'_x)}{n_x - 1} (1 - f_x) \qquad \text{and} \qquad \sum_x l_x^2 \frac{p'_x(1 - p'_x)}{n_x - 1} (1 - f_x)$$

respectively, where $f_x = n_x/N_x$, the sampling fraction in each age-group. These estimates are unbiassed since their expected values are $V(G')$ and $V(R')$ respectively. If the number N_x in each age-group is effectively large, f_x may be neglected.

To carry this discussion further, if circumstances are such that a sample as large as $n_1 = \sum_{x=15.5}^{49.5} n_x$ can only be selected, it can be shown that the variance of G' will be a minimum if the age-structure of the sample is such that

$$\frac{n_x}{n_1} = \frac{\sqrt{p_x(1 - p_x) \dfrac{N_x}{N_x - 1}}}{\sum_x \sqrt{p_x(1 - p_x) \dfrac{N_x}{N_x - 1}}} = \frac{\sqrt{p_x(1 - p_x)}}{\sum_x \sqrt{p_x(1 - p_x)}} \text{ (for large } N_x\text{'s)}. \tag{3}$$

This of course is the analogue of the formula for quantitative variates

given by Neyman (1934). To prove the above result, it is necessary to minimize $V(G')$ with respect to the n_x's subject to the restriction $n_1 = \sum_{x=15.5}^{49.5} n_x$. This may be done by the method of Lagrange multipliers. If cost is a practical consideration, and if it costs c units to investigate every female sampled, and if monetary resources are limited to a units, then a/c females can only be sampled. The age-structure of the sample for which $V(G')$ is a minimum, subject to the cost restriction, will be given by substituting a/c for n_1 in (3). Similarly under the same, size-, or cost-restrictions, $n_2 = \sum_x n_x$ the variance of R' will be a minimum if the allocation of sampling units among age-groups is such that

$$\frac{n_x}{n_2} = \frac{l_x\sqrt{p_x(1-p_x)\dfrac{N_x}{N_x-1}}}{\sum\limits_x l_x\sqrt{p_x(1-p_x)\dfrac{N_x}{N_x-1}}} = \frac{l_x\sqrt{p_x(1-p_x)}}{\sum\limits_x l_x\sqrt{p_x(1-p_x)}} \tag{4}$$

(for large N_x's).

If the problem is considered from the point of view of finding the optimum age-structure of a sample, subject to the restriction that the variance shall be a fixed quantity, the solutions obtained in respect of either G or R are identical with those obtained above.

We may now consider the implications of equations (1), (2), (3) and (4):

(i) It must be admitted that there is no exact solution to the problem of determining sample size since the p_x's are themselves the quantities we are attempting to estimate at the time period in question. However, to obtain a workable solution, the p_x values of a previous time period, not very far off, may be used. In the discussion which follows we shall bear this in mind.

(ii) Consider the problem of sampling so as to have either the estimate of G or R with minimum variance. If an investigator wishes to estimate both these rates from the data of a single sample, a decision will have to be made before the sample units are allocated among the various age-groups, as to whether G' or R' is required with minimum variance, for, an inspection of (3) and (4) shows that the age-structure of a sample of a given size which confers the former with a minimum variance does not[1] do similarly to the latter and vice versa. The fol-

[1]The only trivial case which is an exception to this statement is when l_x = a constant for all age-groups under reference. This implies a state of affairs in which all females between 15 and 50 years have equal chances of survival. In the present condition of mankind this possibility could not be hoped for.

lowing question almost suggests itself: how may control be exercised over the estimate which is not given a minimum variance? We now proceed to answer this question. If the sample is designed to give R' a minimum variance, then its age-structure will be as given in (4), so that in this situation, the variance of the estimate of G will be given by

$$V_G = \frac{1}{n}\left\{\sum_x l_x\sqrt{p_x(1-p_x)\frac{N_x}{N_x-1}}\right\}\left\{\sum_x \frac{1}{l_x}\sqrt{p_x(1-p_x)\frac{N_x}{N_x-1}}\right\}$$

$$-\sum_x \frac{p_x(1-p_x)}{N_x-1} \equiv \frac{1}{n}W_1W_2 - W$$

where n is the size of the sample and the symbols W_1 , W_2 and W, (which have been introduced for the sake of brevity), may be identified in the order presented in the identity. On the other hand, had the age-structure of the sample been according as given by (3), the estimate of G would have had a minimum variance of

$$V_{Gm} = \frac{1}{n}\left\{\sum_x \sqrt{p_x(1-p_x)\frac{N_x}{N_x-1}}\right\}^2 - \sum_x \frac{p_x(1-p_x)}{N_x-1} \equiv \frac{1}{n}W_3^2 - W;$$

whence the relative efficiency of this estimate would be given by

$$E_G = (W_3^2 - nW)/(W_1W_2 - nW), \tag{5a}$$

and, in view of the fact that $W \to 0$, for large N_x's, its limiting value will be given by

$$E_{GL} = \{\sum_x \sqrt{p_x(1-p_x)}\}^2/\{\sum_x l_x\sqrt{p_x(1-p_x)}\}$$

$$\cdot\left\{\sum_x \frac{1}{l_x}\sqrt{p_x(1-p_x)}\right\}. \tag{5b}$$

Clearly E_G cannot exceed E_{GL} so that this limiting relative efficiency, which is independent of sample size, can only be attained in the case of samples from effectively large populations. We also have

$$V_G - V_{Gm} = (W_1W_2 - W_3^2)/n$$

$$= \frac{1}{n}\sum_{x\neq x'} \frac{(l_x - l_{x'})^2}{l_x l_{x'}}\sqrt{p_xp_{x'}(1-p_x)(1-p_{x'})} \tag{6}$$

$$\left(\text{if each } \frac{N_x}{N_x-1} \text{ is taken as } 1\right).$$

The expression $W_1W_2 - W_3^2$ is a function of p, l and N, and on the assumption that some prior knowledge of these quantities exists, suffi-

cient to obtain an approximation, it may be treated as a known constant. It follows that $\delta = V_G - V_{Gm}$, the deviation from the minimum variance, is inversely proportional to the sample size n. Hence for any given δ, the sample size n can be determined approximately. Its accuracy of course will depend on how closely $W_1W_2 - W_3^2$ approximates to its true value. We see that if these n sample units are allocated among the age-groups in the way indicated by (4), R' may be expected to have a minimum variance, and simultaneously, G' may also have a variance differing somewhat closely from its ideal or possible minimum variance. Hence some degree of control over G' has been achieved. This approach has been adopted in view of the fact that demographers attach more importance to the net reproduction rate. With the alternative approach, when G' is given a minimum variance, the relative efficiency of the estimate of R, and the deviation of the variance of the estimate of R from its ideal minimum variance are given by

$$E_R = (W_1^2 - nW_4)/(W_3W_5 - nW_4), \tag{7}$$

and

$$V_R - V_{Rm} = (W_3W_5 - W_1^2)/n$$

$$= \frac{1}{n} \sum_{x \neq x'} (l_x - l_{x'})^2 \sqrt{p_x p_{x'}(1 - p_x)(1 - p_{x'})} \tag{8}$$

$$\left(\text{if each } \frac{N_x}{N_x - 1} \text{ is taken as } 1\right)$$

where

$$W_4 = \sum_x l_x p_x (1 - p_x)/(N_x - 1),$$

$$W_5 = \sum_x l_x^2 \sqrt{p_x(1 - p_x)\frac{N_x}{N_x - 1}},$$

and n, W_1 and W_3 are as defined above.

(iii) In the alternative problem, when the allowable amount of variance is fixed in advance, the investigator who wishes to estimate both G and R from the same data will again encounter the problem of controlling the estimate whose variance has been left unfixed. Suppose for example that the variance of R' is fixed at v. Remembering that the age-structure of this sample will be given according to (4), we have

$$v = \frac{1}{n} W_1^2 - W_4 . \tag{9}$$

Since the conditions which give rise to equation (6) are formally equivalent to those in the present problem, we have

$$\delta = (W_1W_2 - W_3^2)/n$$

By eliminating n between (9) and the above equation, the relation between v and δ is seen to be

$$\delta = \frac{W_1W_2 - W_3^2}{W_1^2} v + \frac{W_4(W_1W_2 - W_3^2)}{W_1^2} . \qquad (10)$$

And on the assumption that the expressions involving the W's can be approximated from prior information, (10) may be used to make numerical adjustments between the deviation from the ideal minimum variance of G', and, the "fixed" variance of R'. After such adjustments are made n can be determined by substituting the value of the "fixed" v in (9). Also since the formulae for the allocation of sampling units in both problems are identical, the relative efficiency of the estimate of G will be as given in equation (5a). Hence we see that some control over the unfixed variance can also be exercised. Similar formulae may be developed for the case when the variance of G' is "fixed" and that of R' unfixed.

(iv) Lastly it may be said that the frame of this sampling system, in which the unit of sampling and the unit of analysis are identical, is perhaps not practicable for large populations of the order of millions. This is true. But for the case of small ethnic groups or minority populations, living within compact geographical areas and in whose prospects of biological survival we may be interested, these methods may be suitable. These populations may number anything between five and fifty thousand. It is also for this reason that the above theory has been developed with a finite approach.

3. ILLUSTRATIVE EXAMPLE

An example to illustrate the above theory on the optimum allocation of sampling units in the estimation of the gross reproduction rate now follows. The data which is presented in columns (i), (ii) and (iii) of the table is drawn from a report on a recent census[2] of the Eurasian

[2]This census, which was sponsored by the Anglo-Burman Union, was carried out by the writer with the assistance of a group of honorary workers in 1949. The writer is grateful to H. A. Faria and W. E. Foster for their generous help in the work of enumeration. Due to the conditions of civil war around Rangoon, the census could not be completed. The Eurasian community numbered about 6,200 at that time. The data on fertility rates is based on 81% of the total female population in the reproductive group. There was also one other defect. It was neglected to ask whether female daughters of age less than 12 months had died in the 12-month period prior to the census. For those age-groups where this defect is present, it is possible that fertility rates may tend to be underestimated, unless compensated by a systematic tendency to select households containing mothers whose female infants survived the first year of life. However, I am of the opinion that this defect is not serious since infantile mortality among Eurasians is not high.

people in Rangoon, which so far has not been published. (Since survival rates of Eurasian women are not known, it will not be possible to give illustrative examples on the problems of sample size, when both the net and the gross reproduction rates are required to be estimated from the data of a single sample.)

TABLE

Age-group (i)	Number of females (ii)	Female infants below one year (iii)	Age-specific fertility rate p_x (iv)	$\sqrt{p_x(1-p_x)}$ (v)	$\frac{\sqrt{p_x(1-p_x)}}{\sum_x \sqrt{p_x(1-p_x)}}$ (vi)
17–18	47	2	.042553	.20185	.02925
18–19	62	0		.20185[1]	.02925
19–20	57	3	.052632	.22330	.03236
20–21	66	7	.106061	.30792	.04462
21–22	60	2	.033333	.17950	.02601
22–23	57	5	.087719	.28289	.04100
23–24	52	4	.076923	.26648	.03862
24–25	46	4	.086957	.28231	.04091
25–26	51	5	.098039	.29918	.04336
26–27	48	8	.166667	.36268	.05259
27–28	38	2	.052632	.22330	.03236
28–29	57	2	.035088	.18400	.02666
29–30	57	5	.087719	.28289	.04100
30–31	52	4	.076923	.26669	.03865
31–32	25	1	.040000	.19596	.02840
32–33	41	4	.097591	.29676	.04301
33–34	30	2	.066667	.24944	.03615
34–35	35	1	.028581	.16667	.02415
35–36	37	1	.027027	.16216	.02350
36–37	36	2	.055556	.22906	.03319
37–38	28	1	.035714	.18558	.02689
38–39	31	0		.18558[2]	.02689
39–40	34	0		.18558	.02689
40–41	44	2	.045455	.20830	.03019
41–42	21	0		.20830	.03019
42–43	26	0		.20830	.03019
43–44	20	0		.20830[3]	.03019
44–45	23	0		.20830	.03019
45–46	39	0		.20830	.03019
46–47	18	1	.055556	.22906	.03319
			$G = 1.455393$	6.90049	1.00004

[1]Token figure, obtained on the assumption that $p_{18.5} = p_{17.5} = .042553$
[2]Token figures, obtained on the assumption that $p_{38.5} = p_{36.5} = p_{37.5} = .035714$
[3]Token figures, obtained on the assumption that $p_{40.5} = p_{41.5} = \cdots = p_{46.5} = .045455$

In column (iv) of the table, the age-specific fertility rates of Eurasian women are given. If these rates are added up, a figure which will somewhat approximate to the true gross reproduction rate will be obtained.

If column (iii) of the table is inspected, it will be seen that no births are shown against women in certain age-groups. It is difficult to decide whether fertility rates for each of these age-groups should be taken as zero, since it was neglected to ask whether female daughters of age less than 12 months had died in the 12-month period prior to the census. If the procedure of gathering data had been statistically random, then there would be complete justification in taking these rates as zero. However, as the data of an incomplete census cannot be taken to be equivalent to that of a random sample, there would be no such justification. If it is presumed that these rates are zero, and the optimum age-structure of the required sample is determined theoretically by formula (3), then no sample units will be allocated to these groups. In the event of their being not really zero, the sampled gross reproduction rate will be underestimated, and hence biassed. For this reason, these groups should also be sampled. And it would be expedient to assign a token rate to each of these groups. In assigning a token rate to any group, in this problem, the rate for the nearest group junior to the group in question has been chosen. This procedure is almost arbitrary, and is founded on the hope that each of the fertility rates chosen for these groups might be about equal in magnitude to its corresponding "missing" rate.

If the sample units are drawn at random from each age-group according to the proportions prescribed in column (vi) of the table, and second, if the age-specific fertility rates used for the purpose of determining these proportions happen to be just the rates prevailing at the period of time when the sampling is carried out, then the variance of the estimate of G will be a minimum. In regard to this problem, even if a sample survey had been carried out immediately after the census, the second set of conditions would not have been realized, for one reason, because of the fact that "missing" rates had been somewhat arbitrarily chosen. However, selecting the sample units in this way, is definitely better than selecting them without any method, and even if G' does not attain minimum variance, the experimenter will have grounds for belief that its variance is somewhere near the minimum.

4. SAMPLING FROM A UNIVERSE OF LARGE PRIMARY UNITS

As stated above, if the population is of the order of millions and spread over a very wide area, clearly the above sampling procedure

will not be practicable. Hence the following alternative procedure is suggested. As in most sample surveys, suppose the area over which the population is distributed is unambiguously divided up along natural or adminsitrative boundaries into K small-size areas. Since the total area is supposed to be very large, the number of these primary units will also be very large, and each will also contain a large number of persons. Following Mahalanobis, these primary units may be termed grids. It may be noted that in each grid we are only interested in women in the reproductive age-groups, and their female children under one year. For a moment we shall defer general consideration on the question of what may be their best relative sizes, sex or age-composition. The notation adopted here will be substantially the same as that above, except for the addition of the suffix i to identify grids. Let n_{xi} be the number of females in the age-group $(x - \frac{1}{2}, x + \frac{1}{2})$ of the ith grid, $(i = 1, 2, \cdots K)$, who have between them m_{xi} female children under one year. In respect of any given age-group in the universe of grids, it would be prudent to take into account the possibility that n_{xi} and m_{xi} may be correlated. The age-specific fertility rate will be given by

$$p_x = \sum_{i=1}^{K} m_{xi} \Big/ \sum_{i=1}^{K} n_{xi} .$$

This is the population value. The problem here is firstly to estimate each of these p_x's so as to enable us to estimate G and R, and secondly to estimate the sampling errors of these estimates.

Let each grid be assigned an equal probability of selection. Let a random sample of k grids be selected and observe in each, the variates n and m, which have been defined above. An estimate p_x'' of p_x will be given by

$$p_x'' = \sum_{i=1}^{k} m_{xi} \Big/ \sum_{i=1}^{k} n_{xi} = \frac{\overline{m}_x}{\overline{n}_x} .$$

This estimate, which is a ratio of two means, though consistent, is biassed, but Hansen and Hurwitz (1943) are of the opinion that the bias "is usually negligible". Hence the estimates of G and R are as follows:

$$G'' = \sum_x \frac{\overline{m}_x}{\overline{n}_x} \qquad \text{and} \qquad R'' = \sum_x l_x \frac{\overline{m}_x}{\overline{n}_x} .$$

The variance of G'' can be shown to be

$$V(G'') = \sum_x V\left(\frac{\overline{m}_x}{\overline{n}_x}\right)$$

$$= \sum_x \left(\frac{\bar{m}_x}{\bar{n}_x}\right)^2 \left[\frac{V(\bar{m}_x)}{\bar{m}_x^2} - 2\frac{Cov(\bar{m}_x, \bar{n}_x)}{\bar{m}_x\bar{n}_x} + \frac{V(\bar{n}_x)}{\bar{n}_x^2}\right] \tag{11}$$

$$= \sum_x \left(\frac{\bar{m}_x}{\bar{n}_x}\right)^2 \left[\frac{V(m_{xi})}{\bar{m}_x^2} - 2\frac{Cov(m_{xi}, n_{xi})}{\bar{m}_x\bar{n}_x} + \frac{V(n_{xi})}{\bar{n}_x^2}\right] \frac{K-k}{k(K-1)} .$$

The expression $(K - k)/(K - 1)$ accounts for the fact that we are sampling a finite population. We have remarked above on the possibility that m_{xi} and n_{xi} may be correlated. The covariance term in (11) accounts for this possibility. The estimate of $V(G'')$ will be seen to be

$$\frac{(1 - k/K)}{k(k-1)} \sum_x p_x''^2 \left[\frac{1}{\bar{m}_x^2} \sum_{i=1}^{k} (m_{xi} - \bar{m}_x)^2 - \frac{2}{\bar{m}_x\bar{n}_x} \sum_{i=1}^{k} (m_{xi} - \bar{m}_x)(n_{xi} - \bar{n}_x) + \frac{1}{\bar{n}_x^2} \sum_{i=1}^{k} (n_{xi} - \bar{n}_x)^2\right].$$

Correspondingly similar expressions may also be written down for the variance of R''. They will only differ in appearance by the term l_x^2 under the summation sign.

It may be noted that the expression for the variance of G'' given in (11) is only an approximate one. Actually it under-estimates the true variance. But it will suffice if each of the variances under the summation sign are small relative to their corresponding squared means (Yates, 1949). In this connection, Finney (1949), alluding to infinite normal populations, when the two variates comprising the ratio are independent, states that $\bar{n}_x^2$ should be at least 40 times $V(\bar{n}_x)$ so as to be able to construct, without danger of error, 95% confidence limits for p_x [3]. Since we are not attempting to find confidence limits for G, this condition perhaps in a less stringent form, applies also to finite populations. Hence to estimate the standard error of G'' using the expression (11), conditions governing its validity must first be satisfied. To do this the only alternative seems to be to demarcate the grids in a way such that the numbers found in each reproductive age-group, ranging from 15–16 to 49–50, be

(i) as near as constant from grid to grid, and

(ii) sufficiently large, subject only to the restriction that the investigator shall not find the grid too unwieldly to handle, or from the point of view of the planning authority, too costly to operate.

(i) and (ii) ensure that for any given age-group, $V(n_{xi})$ will be controlled at a low figure and $\bar{n}_x$ at a high figure. Hence in large scale

[3] I have given the substance of his statement using my own notation.

sampling, when k will obviously be large, the ratio $V(\bar{n}_x)/\bar{n}_x^2$ will be reduced to a very small value. Also since m_{xi} is possibly correlated with n_{xi} , $V(\bar{m}_x)/\bar{m}_x^2$ may be expected to be indirectly controlled at a small value.

The above statements presuppose that prior knowledge, both on the geographical distribution and age composition of the population, is available. In all developed countries, census records will furnish information bearing on these questions. However, because of population changes, by way of births and deaths, and shifts from one locality to another, which are almost certain to occur, such records may not be completely satisfactory for the purpose in hand. But if a survey is timed to occur soon after a national census, then it may be possible to carry out the above suggestions fairly satisfactorily. If the age composition of the female population is not known, but only the total numbers inhabiting each locality are known, the grids may be demarcated such that the sex ratio in each grid and the numbers included in each grid remain almost constant, and also such that condition (ii) is satisfied. More often than not, this may be the only practical course open to the investigator. By this procedure it is quite possible that the females belonging to various age-groups, may be more or less equally distributed over all the grids.

In this system of sampling, the cost of a survey will depend partly on the number of grids selected, and partly on the physical size and population size of each of the grids included in the sample. Hence if monetary resources are fixed at a certain limit, and if we have some empirical knowledge of average costs per grid, the number of grids to be selected will be determined by this limit.

In closing this section it may be stated that alternative systems of sampling are certainly not lacking. Also, what has been suggested above may not be the best in all circumstances. The writer does not propose discussing these problems since the object here is merely to point out that gross and net reproduction rates may be estimated by methods of statistical sampling. Different systems of sampling have been discussed by various writers in the United States, Britain and India, and any of these systems may be used.

5. NOTE ON p_x

A point arising out of this discussion concerns the interval estimation of p_x , the age-specific fertility rate for female births. Suppose that large samples are being selected from very large populations. If p_x is being estimated by the system of sampling outlined in section 2, then its confidence limits, corresponding to the confidence coefficients 0.95

or 0.99, for certain sample sizes, may be read off from charts prepared by Clopper and Pearson (1934). Also Wilks (1943) has shown that approximate confidence limits of p_x , corresponding to the confidence level $\alpha/100$ may be obtained by solving the following quadratic in p_x :

$$u^2 = (p_x - p_x')^2 \Big/ \left\{\frac{p_x(1 - p_x)}{n_x}\right\}, \tag{12}$$

where u is the normal deviate corresponding to the $\alpha\%$ level of significance. On the other hand if p_x is being estimated by the system of sampling outlined in section 4, its approximate confidence limits may be obtained as follows:

Consider the function $d = \overline{m}_x - p_x\overline{n}_x$,

we have $E(d) = 0$;

and $$V(d) = \frac{K - k}{k(K - 1)} [V(m_{xi}) - 2p_x\, Cov(m_{xi} , n_{xi}) + p_x^2 V(n_{xi})]. \tag{13}$$

If K is effectively large, then the factor $(K - k)/(K - 1)$ may be taken as unity. In large scale sampling, $\overline{m}_x$ and $\overline{n}_x$ will be distributed almost normally. Hence

$$u = (\overline{m}_x - p_x\overline{n}_x) \Big/ \left\{\frac{1}{k} [V(m_{xi}) - 2p_x\, Cov(m_{xi} , n_{xi}) + p_x^2 V(n_{xi})]\right\}^{\frac{1}{2}} \tag{14}$$

will be distributed almost normally with zero mean and unit variance. Hence the confidence limits of p_x , corresponding to any desired confidence level, may be obtained by solving the quadratic (14) for p_x . However, since the population values of the covariance and the variance terms in this equation are unknown, the sample values are the only one available for use. Hence the sample size should be large to preserve the normality of u, and preferably $k > 150$. The principle of this method is due to Fieller (1940). This particular problem is just one aspect of the general problem discussed recently by Finney (1949) in this Journal. It is presented here in the notation used in section 4 to show that it has applications in demography.

REFERENCES

Clopper, C. J. and Pearson, E. S. *Biometrika, 26:* 404–413, 1934.
D. V. G. *J. Roy. Statist. Soc., 110:* 383–384, 1947.
Fieller, E. C. *J. Roy. Statist. Soc. Suppl., 7:* 1–64, 1940.
Finney, D. J. *Biometrics, 5:* 335–337, 1949.
Hurwitz, W. N. and Hansen, M. H. *Ann. Math. Stats., 14:* 336, 1943.
Neyman, J. *J. Roy. Statist. Soc., 97:* 558–606, 1934.
Wilks, S. S. *Mathematical Statistics.* Princeton: 129–130, 1943.
Yates, F. *Sampling Methods for Censuses and Surveys.* London: Charles Griffin and Co. Ltd., 198, 1949.

THE USE OF SYSTEMATIC 5 × 5 SQUARES[1]

ØIVIND NISSEN[2]

SYSTEMATIC SQUARES were once used quite commonly by research workers; one of these, the 5 × 5 square which Fisher[3] (3, page 76) calls the Knut Vik square, is still frequently used in Scandinavia. There are two possible different arrangements of this design, one with treatment *A* below *C*, the other with *A* below *D*.

I					II				
A	*B*	*C*	*D*	*E*	*A*	*B*	*C*	*D*	*E*
D	*E*	*A*	*B*	*C*	*C*	*D*	*E*	*A*	*B*
B	*C*	*D*	*E*	*A*	*E*	*A*	*B*	*C*	*D*
E	*A*	*B*	*C*	*D*	*B*	*C*	*D*	*E*	*A*
C	*D*	*E*	*A*	*B*	*D*	*E*	*A*	*B*	*C*

The statistical analysis of these designs were originally made by adjusting for row and column differences. Since Fisher's analysis of variance technique became known, this technique, which in effect gave the same result as the old method, was commonly used. From the total *SS* the *SS* due to rows, columns, and treatments were deducted, and the rest was regarded as error *SS*. Fisher (l.c.) showed that this method of analysis gave a biased estimate of the error. With a Latin square the calculated error is an unbiased estimate of the true error. If, by use of the proper systematic square, however, a lower true error is obtained, the calculated error will automatically increase. Tedin (4) confirmed this statement by calculations on uniformity trials. He found that the true error variance by use of a Knut Vik square was on the average 91.2 per cent of the error variance for Latin squares. The

[1]Contribution from the Department of Agronomy, The Pennsylvania State College, State College, Pa. Authorized for publication on December 3, 1948, as Paper number 1496 in the Journal Series of the Pennsylvania Agricultural Experiment Station.

[2]Visiting Professor, Department of Agronomy, Pennsylvania State College, on leave from Farm Crop Institute, Agricultural College of Norway, Vollebekk, Norway.

[3]Numbers in parentheses indicate Literature cited.

calculated error variance was found to be 13 per cent greater than the true error variance.

Dorph-Petersen (2), however, has shown that it is possible to calculate an unbiased estimate for the error in a Knut Vik square. For an ordinary Latin square there are two restrictions; each treatment shall appear once in each row and once in each column. Now, if two new restrictions are added, namely that each treatment shall appear once in each diagonal (both ways) then one of the two possible Knut Vik squares will be obtained. All plots with the same number in the following scheme are then to be regarded as belonging to the same diagonal:

1	2	3	4	5
5	1	2	3	4
4	5	1	2	3
3	4	5	1	2
2	3	4	5	1

Since rows, columns and both diagonals are orthogonal to each other and to the treatments, the correct analysis of variance for a Knut Vik square will have this form:

	Degrees of freedom
Rows	4
Columns	4
Diagonals 1st way	4
Diagonals 2nd way	4
Treatments	4
Error	4
	24

However, it is not necessary to carry out the whole analysis in order to find the error. With arrangement *I* (p. 1) it is necessary only to calculate the sums of the plots with the same letter in arrangement *II*. The variance between these sums then becomes the error variance. This procedure utilizes the fact that the design is a completely orthogonal 5×5 square (Fisher, l.c. p. 81). This method of calculating the error in a systematic 5×5 square was proposed by Vik (5) as early as 1924, and has been called "Vik's short-cut method". Dorph-Petersen (l.c.) confirmed these theoretical considerations by comparing *F*-values from uniformity trials with the theoretical *F*-distribution and obtained a satisfying agreement, at least for $P = 0.05$.

The main disadvantage of the systematic square is that there are

only 4 degrees of freedom for error whereas there are 12 in a 5 × 5 Latin square.

According to Cochran & Cox (1, p. 29) the relative efficiency for two designs with the same error variance and with n_1 and n_2 degrees of freedom for error is

$$\frac{(n_1 + 1)(n_2 + 3)}{(n_2 + 1)(n_1 + 3)}$$

When comparing a Knut Vik square ($n_1 = 4$) with a Latin square ($n_2 = 12$) this fraction is $(5 \times 15)/(13 \times 7)$ or 82.4%. As mentioned, Tedin (l.c.) found that the average true error variance for the Knut Vik square was 91.2% of the error variance for the Latin square. For a single experiment the ordinary Latin square is therefore the most efficient. In many cases, there is no special interest in the error of a single experiment. When a series of experiments is conducted in different years and in different localities, it is necessary to determine an average value for the experimental error in order to test the significance of the different interactions (treatment × year, treatment × locality, etc.). In this case, the relative efficiency of the systematic square is much higher than for a single experiment. For a small series of 5 experiments $n_1 = 20$ (Knut Vik square) and $n_2 = 60$ (Latin square) and the fraction $(21 \times 63)/(61 \times 23) = 94.3\%$. If it is assumed that the error variance for the systematic square is 91.2% of the error variance for the Latin square, the systematic square will be the more efficient one. For larger series this advantage of the systematic square will be still more pronounced.

In such cases the important error is the error "between experiments". This error will be the composite of the true variability between experiments and the variability within each experiment. The first part is independent of the design used. If therefore, by using the Knut Vik square instead of the Latin square the later part can be reduced, the error "between experiments" will also be reduced thus increasing the precision of the whole series of experiments. This increased precision is obtained without any increased cost. In fact, by using the Knut Vik square, the cost of planning, handling, and supervising a series of similar experiments will be lower than by using Latin squares, and the chance of making errors will decrease. By means of Vik's short-cut method, the labor involved in the calculation of the error within experiments will also be less than by use of Latin squares. The only disadvantage is that to some extent it is necessary to sacrifice some information from individual experiments. In working with a large series of experiments, the information from a single experiment is, how-

ever, of minor importance and in such cases the use of the Knut Vik square can be recommended. Of course the analysis of data obtained with these squares depends on the assumption of normality; there being only two possible Knut Vik squares, one cannot use the randomization argument to justify F tests as Pitman and Welch have done for ordinary Latin squares. On the other hand, results of Dorph-Petersen in comparing F-values from uniformity trials to the theoretical F distribution provide justification for using F tests.

REFERENCES

1. Cochran, W. G. and Cox, G. M. *Experimental designs*, John Wiley & Sons, Inc., New York, 1950.
2. Dorph-Petersen, K. Fejlberegning paa Forsøg med systematisk Parcelfordeling. *Nordisk Jordbrugsforskning 24:* 140–150, 1942.
3. Fisher, R. A. *The design of experiments.* 3 ed., Oliver & Boyd, London, 1942.
4. Tedin, O. The influence of systematic plot arrangement upon the estimate of error in field experiments. *The Journal of Agricultural Science 21:* 191–208, 1931.
5. Vik, K. Bedømmelse av feilen på forsøksfelter med og uten målestokk. *Meldinger fra Norges Landbrukshøgskole 4:* 129–181, 1924.

THE EFFECT OF SULFAMERAZINE ON THE ERYTHROCYTE AND HEMOGLOBIN CONTENT OF TROUT BLOOD

JAMES S. GUTSELL

U. S. Department of the Interior, Microbiological Laboratory
Fish and Wildlife Service
Kearneysville, W. Va.

INTRODUCTION

THE DISCOVERY, at Leetown in 1945, of the effectiveness of sulfamerazine in the treatment of furunculosis in trout (1, 2), and the consequent widespread use of this drug for this purpose, made it important to learn the effects of this sulfonamide on trout. Accordingly an experiment was run to test the effects of various dosage rates, from approximately the minimum effective rate to one greatly in excess of the highest rate recommended.

The daily dosage rate originally recommended (2) was 8 grams per 100 pounds of fish. Later it was found that a lighter rate was sufficient and that a 4-gram rate was about as effective as the 6-gram rate subsequently recommended (3, 4, 5). The dosage rates used in this experiment were 0, 5, 10 and 15 grams of sulfamerazine per day per 100 pounds of fish.

The duration of treatment originally recommended was three weeks. In this experiment sulfamerazine was administered through more than twice this period.

In the original experiment (1, 2) the species of trout used was the Eastern brook trout, *Salvelinus fontinalis*, a species very susceptible to furunculosis. This experiment, described in part by Gutsell and Snieszko (12), included two other species: the brown trout, *Salmo trutta* or *Salmo fario*, generally less susceptible; and the rainbow trout, *Salmo gairdnerii*, which only occasionally suffers noticeable loss from this disease.

The test using brook and rainbow trout was run in 1946; that using brown trout in 1947. Erythrocyte counts and hemoglobin measurements were made only in 1947 and therefore could include only the brown trout.

With rainbow trout, growth was not affected even by the heaviest dosage. With brook trout, growth was much reduced only by the heavier dosage rates; with brown trout, even by the lightest rate. The reduced growth in brook and brown trout appeared to be due to reduced food consumption due to dislike of the taste of the food containing the drug. In this and in other experiments run at Leetown, there was not any mortality attributable to the use of sulfamerazine.

METHODS AND PROCEDURE

As noted above, the daily sulfamerazine dosage rates were 0, 5, 10 and 15 grams per 100 pounds of fish (0, 11, 22, and 33 grams per 100 kilograms). The drug was administered by being mixed very thoroughly with the food. Except for the 0-rate, all food used in the experiment contained the drug. Treatment was continued through 49 days.

Five pounds of brown trout from one large pool were placed in each of 8 adjacent upper troughs in the hatchery. Each of the 4 treatments was assigned to 2 troughs by random selection, so that the test was run in duplicate and was fully randomized.

The hemoglobin measurements were made after 35 days of treatment; the erythrocyte counts, after 42 days. A Helige hemometer was used for the hemoglobin measurements and a Spencer Bright-line counting chamber for erythrocyte counts. The usual technique for these operations was followed, except that Dr. S. F. Snieszko, who made the measurements and counts, used his *cold method* (Pipettes kept at a low temperature and fish placed in ice-water, to prevent the clotting which is the bane of such work with trout blood). For the erythrocyte counts, a 1–100 dilution was used.

The fish used for the hemoglobin measurements and red cell counts were taken at random from each trough by means of a dip net.

For hemoglobin measurements, 10 fish were taken from each trough, or 20 fish for each treatment. For erythrocyte counts, only half this number of fish were used, but two counting chambers (*a* and *b*) were filled from each fish. Thus there were 20 erythrocyte counts and 20 hemoglobin measurements for each treatment.

THE DATA AND THEIR ANALYSIS

The average hemoglobin and erythrocyte contents are shown in Figure 1; the hemoglobin measurements and erythrocyte counts in Tables 1 and 3. Hemoglobin is expressed as grams per 100 milliliters of blood. Each erythrocyte count in Table 3 is the sum of the counts for 5 secondary squares. Because the dilution was 1–100, each count

in Table 3 (or each average of such counts) should be divided by 2 before the 4 zeros (0, 000) are added to give the number of erythrocytes per cubic millimeter of blood.[1]

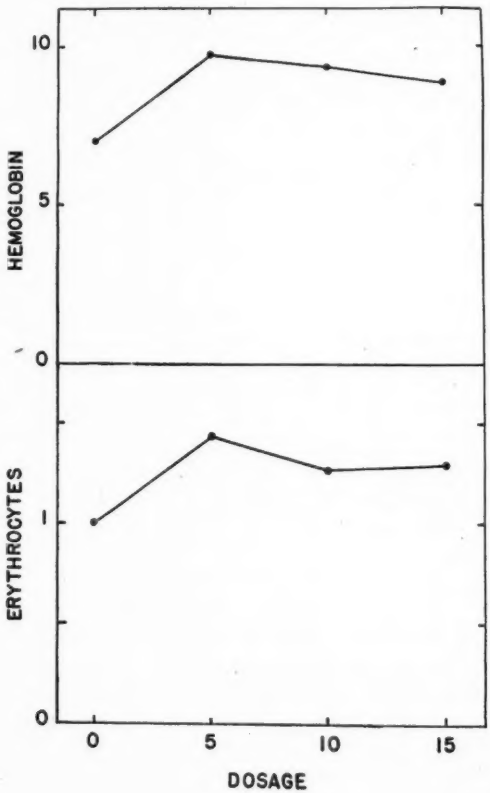


FIGURE 1.
Hemoglobin (grams per 100 ml.) and erythrocytes (millions per cu. mm.) in the blood of brown trout after 35 (hemoglobin) and 42 (erythrocytes) days of treatment at the daily dosage rates of 0, 5, 10 and 15 grams of sulfamerazine per 100 pounds of fish. Each point represents the average of 20 determinations.

[1]Each count is the number of red cells in 5 secondary squares of the counting chamber. The volume of the diluted blood in the 5 secondary squares is 1/50 of a cubic millimeter. Therefore, the estimated number of red cells in a cubic millimeter of the diluted blood is 50 times the count. If the dilution was 1/100, the estimated number of erythrocytes per cubic millimeter of blood is 5,000 (100 × 50) times the count. This multiplication is most conveniently done by dividing the count by 2 and adding 4 zeros (0,000; equivalent to multiplying by 10,000).

TABLE 1
MEASUREMENTS OF HEMOGLOBIN (GRAMS PER 100 ML.) IN THE BLOOD OF BROWN TROUT AFTER 35 DAYS OF TREATMENT AT THE DAILY RATES OF 0, 5, 10 AND 15 GRAMS OF SULFAMERAZINE PER 100 POUNDS OF FISH. TWO TROUGHS, *A* AND *B*, FOR EACH TREATMENT; 1 MEASUREMENT FROM EACH OF 10 FISH FROM EACH TROUGH.

Fish no.	Treatment 0		Treatment 5		Treatment 10		Treatment 15	
	A	*B*	*A*	*B*	*A*	*B*	*A*	*B*
1	6.7	7.0	9.9	9.9	10.4	9.9	9.3	11.0
2	7.8	7.8	8.4	9.6	8.1	9.6	9.3	9.3
3	5.5	6.8	10.4	10.2	10.6	10.2	7.2	11.0
4	8.4	7.0	9.3	10.4	8.7	10.4	7.8	9.0
5	7.0	7.5	10.7	11.3	10.7	11.3	9.3	8.4
6	7.8	6.5	11.9	9.1	9.1	10.9	10.2	8.4
7	8.6	5.8	7.1	9.0	8.8	8.0	8.7	6.8
8	7.4	7.1	6.4	10.6	8.1	10.2	8.6	7.2
9	5.8	6.5	8.6	11.7	7.8	6.1	9.3	8.1
10	7.0	5.5	10.6	9.6	8.0	10.7	7.2	11.0
Trough sums	72.0	67.5	93.3	101.4	90.3	97.3	86.9	90.2
Treatment sums	139.5		194.7		187.6		177.1	
means	6.98		9.74		9.38		8.86	
Total	698.9							

Analyses of variance of the data are given in Tables 2 (hemoglobin) and 4 (erythrocytes). The higher hemoglobin and erythrocyte contents of the blood of fish given sulfamerazine are found to be highly significant (F for treatments above the 0.1 percent probability points). For the erythrocyte data, this applies not only to the counts but also to the square root and logarithmic transformations (see bottom of Table 4). Neither transformation changed the picture.

In the breakdown of the three degrees of freedom for treatment, for each degree F is above the 5 percent probability point, and for all but one degree (third degree, hemoglobin) above the one precent probability point.

To get at this another way, the values for the 5-gram rate were compared with those for each of the others. For hemoglobin the comparisons were as follows: 5-gram rate *vs.* 0-rate, $F = 49.11$; 5-gram rate *vs.* 10-gram rate, $F = 0.81$; 5-gram rate *vs.* 15-gram rate, $F = 4.99$

TABLE 2
ANALYSIS OF VARIANCE OF HEMOGLOBIN DATA.

Source of variation	Degrees of freedom	Sums of squares	Mean squares	*F*	Probability Points		
					5%	1%	0.1%
Total	79	209.48					
Treatments .	3	90.56	30.19	19.48	2.74	4.08	6.09
(1st°) Linear .	1	27.93	27.93	18.02	3.98	7.01	11.87
(2nd°) Quadratic . . .	1	53.96	53.96	34.81	3.98	7.01	11.87
(3rd°) Cubic .	1	8.67	8.67	5.59	3.98	7.01	
Between troughs treated alike	4	7.29	1.82	1.17	2.50		
Between fish in same trough	72	111.64	1.55				

TABLE 3
ERYTHROCYTE COUNTS[1] (5 SECONDARY SQUARES; DILUTION, 1–100) FROM BROWN TROUT AFTER 42 DAYS OF TREATMENT AT DAILY RATES OF 0, 5, 10 AND 15 GRAMS OF SULFAMERAZINE PER 100 POUNDS OF FISH. TWO TROUGHS, *A* AND *B*, FOR EACH TREATMENT; 5 FISH FROM EACH TROUGH; 2 SLIDES, *a* AND *b*, FROM EACH FISH. TREATMENT MEANS SHOWN (IN PARENTHESES) AS NUMBERS PER CU. MM.

Fish no.	Treatment 0				Treatment 5				Treatment 10				Treatment 15			
	A		*B*		*A*		*B*		*A*		*B*		*A*		*B*	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
1	213	230	166	157	296	319	310	309	339	322	196	232	278	212	287	280
2	253	231	206	185	278	258	241	270	282	285	205	186	275	311	221	243
3	195	164	245	250	345	307	272	311	236	262	252	274	186	158	331	309
4	193	203	213	181	322	372	254	237	252	209	245	216	301	281	231	244
5	191	195	198	169	248	274	266	275	263	296	249	260	223	246	292	295
Column sums .	1045	1023	1028	942	1489	1530	1343	1402	1372	1374	1147	1168	1263	1208	1362	1371
Trough sums . .	2068		1970		3019		2745		2746		2315		2471		2733	
Treatment sums	4038				5764				5061				5204			
Treatment means	201.9 (1,009,500)				288.2 (1,441,000)				253.05 (1,265,250)				260.2 (1,301,000)			
Total	20,067															

[1]To convert counts to numbers per cubic millimeter, divide by 2 and add 0,000; thus 230/2 = 115; 115 plus 0,000 = 1,150,000. (See footnote to text).

(F .05 = 3.98). For erythrocytes: 5-gram rate *vs.* 0-rate, F = 30.45; 5-gram rate *vs.* 10-gram rate, F = 5.05; 5-gram rate *vs.* 15-gram rate, F = 3.21 (F .1 = 2.88, F .05 = 4.15).

In a t test, a comparison of the 5-gram rate with the 15-gram rate (hemoglobin) gave t = 2.078 (t .05 = 2.025).

TABLE 4
ANALYSIS OF VARIANCE OF ERYTHROCYTE COUNTS

Source of Variation	Degrees of freedom	Sums of squares	Mean squares	F[1]	Probability points		
					5%	1%	0.1%
Total	79	186,948.89					
Treatments	3	77,667.74	25,889.25	10.58	2.90	4.46	6.96
(1st°) Linear	1	19,530.06	19,530.06	7.98	4.15	7.50	13.15
(2nd°) Quadratic	1	31,323.61	31,323.61	12.81	4.15	7.50	13.15
(3rd°) Cubic	1	26,814.06	26,814.06	10.96	4.15	7.50	13.15
Between troughs treated alike	4	16,954.25	4,238.56	1.73	2.67		
Between fish in same trough	32	78,267.40	2,445.86	6.93	1.74	2.20	2.6
Between samples from same fish	40	14,059.50	351.49				

[1]After square root transformation, the values found for F for treatments, troughs and fish were 11.77, 1.82, 4.37; after logarithmic transformation, 11.01, 1.61 and 4.01.

DISCUSSION

The analyses of the data strongly indicate that the higher hemoglobin and erythrocyte concentrations in the blood of the brown trout on the diets including sulfamerazine is causally connected with the presence of sulfamerazine in the diet. There is considerable indication that the higher content associated with the 5-gram drug rate is significant (see Tables 2 and 4, individual degrees of freedom for treatment, also reference in the text to tests of significance), that the 5-gram rate truly is superior in this respect (perhaps near the optimum rate) under the conditions of the experiment.

An important consideration is the relationship of the hemoglobin and red cell concentrations associated with the various treatments, including the 0-rate, to the normal concentrations. What are the normal concentrations? Schlicher (6, quoted by Wunder, 7) listed 1,140,000 erythrocytes per cubic millimeter of blood for brown trout, sensibly the same as for rainbow trout (1,100,000). Phillips (8, Fig. 1) found above 1,500,000 erythrocytes per cubic millimeter of blood from brook trout before they were put on a synthetic diet and after they were given beef

liver to the extent of half of the diet. Tunison *et al*, (9) present these same findings and others showing brook trout on a straight beef liver diet with 1,176,666 red cells per cubic millimeter. Tunison *et al*, (10) reported the following yearly averages for trout on a diet of beef liver (50 percent), pork spleen (48 percent) and salt (2 percent): rainbow, 1,220,000; brook, 1,032,000; brown, 766,000 erythrocytes per cubic millimeter of blood. Phillips *et al*, (11) found the following number of red cells per cubic millimeter of blood after the trout had been 22 weeks on a diet of beef liver (98 percent) and salt (2 percent): rainbow, 1,165,000; brook, 1,186,000; brown 1,099,000. On a diet of beef spleen (98 percent) and salt, the numbers were very similar: rainbow, 1,164,000; brook, 1,181,000; and brown, 1,116,000. Schlicher (see above) gave the hemoglobin level for brown trout as 60 percent (presumably of the normal for human blood or .6(14.5) = 8.7 grams per 100 milliliters).

If it were not for the one high level, shown by Phillips (8, Fig. 1) and Tunison *et al*, (9, Chart 11), a fair assumption from the above references would be that for brook, rainbow and brown trout the normal is close to 1,000,000—1,200,000 per cubic millimeter. This is above the average for the "control" trout in this experiment, below the average for the others. If about 1,500,000 (8, 9) is normal, only the trout treated at the 5-gram rate showed an average approaching the normal red cell concentration.

Another important consideration is the probable sulfamerazine concentrations or levels obtained with the various treatment rates. For brown trout, data are available for approximately the 5-gram rate. Snieszko and Friddle (unpublished) used the daily rate of 0.1 gram of sulfamerazine per kilogram of trout (4.54 grams per 100 pounds) with this species. After a week of treatment the following levels were found: blood, 64; liver, 24; kidney, 24; muscle, 11 milligrams percent. Thereafter, the following more moderate levels (approximate) prevailed: blood, liver and kidney, 8; muscle 5. This was the only rate studied with brown trout; but with brook trout, 0.1, 0.2 and 0.3 grams per day per kilogram of trout (roughly corresponding with the 5, 10, and 15 grams per 100 pounds used in this experiment) were investigated. The following levels approximate those which prevailed after three days in gutted fish and presumably are close to the levels for the muscle: 0.1, 3; 0.2, 6; and 0.3, 7 milligrams percent. These findings indicate that the 5-gram rate is unusually effective with brown trout in producing satisfactory therapeutic sulfamerazine levels, and that the high hemoglobin and erythrocyte levels which followed prolonged treatment with sulfamerazine were associated with fairly high sulfamerazine levels in the blood and tissues.

That the sulfamerazine treatment, almost certainly accompanied by high sulfonamide blood levels, did not reduce the hemoglobin and erythrocyte concentrations is notably important; that these concentrations were increased is of great interest and of possible importance. The considerably higher concentration associated with the use of one of the three dosage rates is of additional interest.

Because prolonged sulfamerazine treatments at high dosage rates had not been accompanied or followed by any trout mortality attributable to the drug (2, 5, 12), the lack of a reduction in the concentration of hemoglobin and erythrocytes was not altogether surprising; but the increase was a great surprise and still is a puzzle. Gutsell and Snieszko (12) offered the following possible explanations: 1, some infection, apparently not furunculosis (13), causing a moderate anemia was cured or greatly reduced by sulfamerazine therapy; 2, the sulfonamide to some degree inhibited the respiratory enzymes (14) with the effect of a reduction in the oxygen supply which is (15) to increase the hemoglobin and erythrocyte concentrations; 3, sulfamerazine, instead of having an anemic effect, stimulates the production of hemoglobin and erythrocytes in trout. The writer offers a fourth possible explanation, that the diet caused a moderate anemia, and that the sulfamerazine, by reducing the consumption of food (as was observed to be the case), reduced the anemia (and growth, of course).

The clear evidence that continued dosage with sulfamerazine did not reduce the hemoglobin or erythrocyte concentrations explains in some measure why adverse effects, other than poor food consumption (apparently due to dislike of the taste of the drug) and reduced growth, have not been observed at Leetown, either in experiments or in practical application of sulfamerazine therapy for furunculosis.

The author wishes to thank Dr. C. M. Mottley for suggestions for the analysis of these data, especially the erythrocyte counts, and for reading the manuscript. Dr. S. F. Snieszko, as a collaborator in the investigation, made the erythrocyte counts and hemoglobin measurements. Mr. Saufley Friddle helped in the preparation of the diets and in feeding the trout. Messers. E. W. Surber and B. H. Hazen made hatchery troughs available and extended numerous courtesies.

REFERENCES

(1) Gutsell, James S. Sulfa drugs and the treatment of furunculosis in trout. *Science 104:* 85–86, 1946.

(2) Gutsell, James S. The value of certain drugs, especially sulfa drugs, in the treatment of furunculosis in brook trout, *Salvelinus fontinalis. Trans. Am. Fish. Soc. 75* (1945), 186–199, 1948.

(3) Gutsell, James S. Furunculosis and its treatment. *The Progressive Fish-Culturist 9:* 13–20, 1947.

(4) Gutsell, James S. and S. F. Snieszko. Furunculosis in fish; its diagnosis and treatment. *Vet. Med. 43:* 484–486, 1948.

(5) Gutsell, James S. and S. F. Snieszko. Dosage of sulfamerazine in the treatment of furunculosis in brook trout, *Salvelinus fontinalis. Trans. Am. Fish. Soc. 76* (1946), 82–96, 1949.

(6) Schlicher, J. Vergleichend-physiol. Untersuchungen der Blukörperchenzahlen bei Knochenfischen. *Zool. Jahrb. Allg. Zool. 43:* 1927.

(7) Wunder, Wilhelm. *Physiologie der Süsswasserfische Mitteleuropas.* E. Schweizerbart (Erwin Nägele), Stuttgart, 1936.

(8) Phillips, Arthur M., Jr. Meatless diets and anemia. *The Progressive Fish-Culturist 48:* 11–13, 1940.

(9) Tunison, A. V., A. M. Phillips, C. M. McCay, C. R. Mitchell and E. O. Rodgers. The nutrition of trout. *The Cortland Hatchery Report No. 8,* for the year 1939.

(10) Tunison, A. V., D. R. Brockway, H. B. Shaffer, J. M. Maxwell, C. M. McCay, C. E. Palm and D. A. Webster. The nutrition of trout. *Cortland Hatchery Report No. 12,* for the year 1943.

(11) Phillips, Arthur M., Jr., D. R. Brockway, E. O. Rodgers, R. L. Robertson, H. Goodell, J. A. Thompson and H. Willoughby. The nutrition of trout. *Cortland Hatchery Report No. 16,* for the year 1947.

(12) Gutsell, James S. and S. F. Snieszko. Response of brook, rainbow and brown trout to various dosages of sulfamerazine. *Trans. Am. Fish. Soc. 77:* 93–101, 1949.

(13) Field, John B., Lynn L. Gee, C. C. Elvehjem, and Chancey Juday. The blood picture in furunculosis induced by *Bacterium salmonicida* in fish. *Arch. of Biochem. 3:* 277–284, 1946.

(14) Altman, K. I. The effects of sulfonamides on respiratory enzymes. *Jour. Biol. Chem. 166:* 149–160, 1946.

(15) Phillips, Arthur M., Jr., Donald R. Brockway, E. O. Rodgers, M. W. Sullivan, Blendon Cook, J. R. Chipman. The nutrition of trout, *Cortland Hatchery Report No. 15,* for the year 1946.

SUR UNE APPLICATION DE LA STATISTIQUE MATHÉMATIQUE À LA BIOLOGIE

MAURICE FRÉCHET

à Paris

Rappel.—Nous avons consacré ces dernières années trois mémoires[1] à l'étude des éléments aléatoires de nature quelconque, en vue de généraliser la théorie des *nombres* aléatoires, objet essentiel, jusqu'à présent, du calcul des Probabilités. Dans cette étude, nous avons défini de plusieurs manières les positions typiques d'un élément aléatoire de nature quelconque, par extension de la notion de "valeur typique" (valeur moyenne, équiprobable, dominante, etc.) d'un *nombre* aléatoire. Si, par exemple, on a défini une "distance" (a, b) de deux quelconques, a, b, des déterminations que peut prendre l'élément aléatoire X, on pourra opérer ainsi: Soit $\theta(a)$ la racine d'ordre α de la valeur moyenne de la puissance $\alpha(>0)$ de la distance (X, a)

$$\theta(a) = \sqrt[\alpha]{\mathfrak{M}(X, a)^{\alpha}}.$$

S'il existe une position γ_α de l'élément certain a, tel que l'écart moyen, $\theta(a)$, d'ordre α de X avec a, atteigne sa borne inférieure quand a devient γ_α, nous appellerons γ_α une *position typique d'ordre* α de X. Ce sera une position *moyenne* pour $\alpha = 2$, une position *équiprobable* pour $\alpha = 1$.

Quand, par exemple, l'élément X est une fonction $X(t)$ choisie au hasard parmi une famille F de fonctions numériques continues $f(t)$ sur $(0, 1)$ et si

$$(f, g) = \max_{0 \le t \le 1} | f(t) - g(t) |$$

la position moyenne de cet élément, $\overline{X}(t)$, est une fonction continue qui *n'a pas* une définition particulière pour chaque valeur de t mais

[1] M. Fréchet, Les éléments aléatoires de nature quelconque dans un espace distancié, *Annales de l'Institut H. Poincaré*, *14* (1948), p. 215–310; L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et ses applications à la moyenne d'un élément aléatoire de nature quelconque, *Revue Scientifique*, *88* (1944), p. 483–485; Positions typiques d'un élément aléatoire de nature quelconque, *Annales École Normale Supérieure*, *65* (1948), p. 211–237.

qui est définie comme un bloc. Elle doit être distinguée de la fonction $\overline{X(t)}$ qui, pour chaque valeur de t, est la moyenne au sens classique du *nombre* aléatoire $X(t)$. Et nous avons montré par un exemple que ces deux définitions peuvent conduire à des fonctions numériquement différentes.

D'une façon générale, quand un élément aléatoire Y est déterminé par la connaissance de certaines caractéristiques numériques $Y_1, Y_2, \cdots$ il n'y a pas de raison pour que notre définition conduise à prendre pour élément moyen $\overline{Y}$ un élément déterminé par des caractéristiques $y_1 = \overline{Y_1}, y_2 = \overline{Y_2}, \cdots$ respectivement égales aux moyennes de Y_1, de $Y_2, \cdots$.

Application à l'homme moyen.—Ayant eu tout récemment l'occasion d'entendre renouveler les critiques—d'ailleurs fondées—adressées maintes fois à la conception, due à QUETELET, *de l'homme moyen*, il nous a paru qu'on pouvait les éviter *en modifiant* la définition de l'homme moyen conformément à la dernière remarque ci-dessus.

En effet, d'après QUETELET, l'homme moyen (d'une population donnée) serait par définition celui dont chaque caractéristique (poids, taille, etc.) serait mesurée par la moyenne des valeurs de cette caractéristique (dans la population donnée).

Joseph BERTRAND a fait observer qu'une telle définition pourrait conduire à un résultat impossible ou anormal, que, par exemple, le poids moyen pourrait n'être le poids d'aucun individu ayant la taille moyenne ou (plus vraisemblablement) être beaucoup plus fort que les poids de la plupart des individus ayant la taille moyenne.

On échappera à cette objection de principe en appliquant à une population humaine notre définition générale ci-dessus. Nous l'exposerons sous une forme un peu modifiée. Supposons qu'on puisse, connaissant deux individus I, I' et leurs caractéristiques $p, t, \cdots, p', t', \cdots$ estimer numériquement leur dissemblance au moyen d'un nombre (I, I') qui sera une fonction déterminée des différences en valeurs absolues $|p - p'|, |t - t'|. \cdots$ (On pourra, par exemple, prendre pour (I, I') une valeur typique de l'ensemble des nombres $|p - p'|, |t - t'|, \cdots$). Alors, on pourra définir un degré numérique de dissemblance de I avec l'ensemble de la population P au moyen d'une valeur typique $[I, P]$ de l'ensemble des valeurs de (I, I') pour I fixe, I' variable pour P.

Ceci étant, l'homme typique (de la popuiation) sera celui, J, dont le degré de dissemblance $[J, P]$ avec la population sera le plus faible, c'est-à-dire tel que $[J, P] \leq [I, P]$ quel que soit l'individu I de la population P.

Remarque: Il arrivera généralement que l'homme moyen (ou typique)

selon la nouvelle définition sera exceptionnel par quelque trait. Il échappera ainsi à la critique de BERTRAND selon laquelle l'homme moyen serait "médiocre en tout".

Si l'on enlève à la critique de BERTRAND sa pointe (car ici "médiocre" n'est que le sens péjoratif de "moyen"), on peut cependant se demander si elle concerne vraiment un défaut et s'il ne conviendrait pas que l'homme moyen soit,—sinon moyen en tout comme le voulait QUETELET ce qui peut être impossible—, du moins qu'il soit aussi près que possible d'être moyen en tout. C'est ce qui nous a amené, indépendamment, Mme DEFRISE (à qui j'avais communiqué la définition ci-dessus) et moi-même à définir autrement l'homme moyen: ce serait ainsi celui des membres de la population considérée dont l'ensemble des caractéristiques diffère le moins de l'ensemble des caractéristiques moyennes. La différence globale pourra être mesurée par une valeur typique de l'ensemble des valeurs absolues des différences entre chaque caractéristique de l'individu considéré et la caractéristique moyenne correspondante. On pourra appeler l'homme ainsi choisi, l'homme le moins éloigné de la moyenne ou plus brièvement l'homme le plus moyen. On définirait, de même, l'homme le plus médian, ··· l'homme le plus typique.

Nous développerons ailleurs plus en détail ces diverses définitions de l'homme typique par exemple en ce qui concerne le repérage numérique des caractéristiques.

Détermination de l'espèce.—Nous avions remis en Belgique puis envoyé de Paris deux notes successives sur cette conception de l'homme moyen ou typique (*pure application* de nos conceptions générales antérieures, rappelées ci-dessus), lorsque nous avons lu à notre retour à Paris, un article, court mais très intéressant, par M. S. KULCZYŃSKI sous le titre "Sur quelques applications des notions topologiques à la biologie".[2] On verra tout de suite la relation de la proposition de cet auteur avec la nôtre, si on la résume en deux points: 1°. On représente chaque individu I d'une population animale ou végétale par un certain nombre de valeurs caractéristiques P, T, ··· qui peuvent être considérés comme ses coordonnées dans dans un certain espace. L'ensemble des points correspondant aux individus de la population forme un nuage; 2° Chaque "point d'accumulation" de ce nuage détermine un individu de la population qui définira *une espèce*.

Ainsi nous voyons que chaque espèce est représentée par un individu pris dans l'espèce elle-même, ce qui le fait échapper à l'objection de BERTRAND, et qui constitue ainsi une amélioration de la définition de

[2] *Colloquium Math.*, *1* (1948), p. 182–183.

QUETELET. Bien que l'auteur considère qu'il y a là une application de la topologie, nous croyons qu'il s'agit plutôt d'une application de la statistique mathématique. L'auteur laissant indéterminée la définition de ce qu'il appelle un point d'accumulation, nous pensons que son interprêtation la plus satisfaisante serait une position où la densité statistique atteint un maximum local' ou mieux comme une position typique au sens de la statistique mathématique sous la forme généralisée ci-dessus pour l'un des nauges de points séparés par des régions à densité statistique nulle on minime.

Généralisations.—Il nous a paru en tout cas intéressant de signaler la convergence (en même temps que l'indépendance) des conceptions de M. KULCZYŃSKI et des nôtres. Et aussi de faire observer que ces conceptions de l'individu représentatif d'une espèce ou d'une population d'êtres vivants peuvent être étendues au cas beaucoup plus général d'une population d'éléments de nature quelconque. C'est ainsi que nous rassemblons actuellement les éléments statistiques permettant de définir une ville typique (d'un ensemble assez homogène de villes d'importances comparables).

Dans les mémoires cités plus haut[3] nous avons d'ailleurs aussi étudié d'autres secteurs (comme, par exemple, l'étude des formes variées de convergence stochastique) de la généralisation de la théorie des *nombres* aléatoires à celle des éléments aléatoires de nature quelconque.

Addendum.—Lors de notre exposé sur ce qui précède, devant la nouvelle *Société française de Biométrie*, le 17 mai 1949, un de nos auditeurs, M. DELAPORTE, nous a signalé qu'il avait, indépendamment de M. KULCZYŃSKI, exposé des idées analogues à celles de ce dernier, en 1946, aux pages 14–23 du tome 8 du périodique édité par la Société de *Biotypologie* (et même, auparavant, dans une communication non publiée faite au *Congrès de l'Association française pour l'Avancement des Sciences* en Octobre 1945). On trouvera une mise au point récente de ses idées dans son article "Bases théoriques de la détermination des types", dans la Revue *Sciences*, Paris, *76*(1949), p. 110–119.

Nous nous étions rencontré avec lui dans notre communication en remplaçant le point d'accumulation de M. KULCZYŃSKI déterminant l'espèce (ou le type) par un point de densité maximum locale qu'on peut appeler point dominant local.

Bien entendu, il n'y a pas lieu, généralement, de considérer *tout* point où la densité de probabilité (ou, en pratique, la densité de

[3](Cf. note [1]). Nous avons résumé une partie de ces mémoires: en français dans le *Journal de la Société de Statistique de Paris, 88* (1947), p. 410–420 et en anglais dans *Mathematics Magazine, 22* (1948–49), p. 1–12.

fréquence) est plus grande que pour les points voisins, comme définissant un type ou une espèce. Il pourra arriver que ce soit simplement un point remarquable de la répartition de la population relativement à ses caractères ou simplement un effet du hasard. Il faudrait s'assurer si l'on peut répartir la population en plusieurs lots dont les nuages de points représentatifs correspondants s'amassent nettement chacun autour d'une position déterminée. Et celle-ci ne sera pas nécessairement un point dominant mais un point moyen ou médian, ··· ou en général typique au sens précisé plus haut. La distinction en types sera plus nette si les nauges correspondants sont complètement séparés les uns des autres. Mais cela n'arrivera guère que pour des types très différents ou dans les cas où la population totale est peu nombreuse, cas où l'existence de nuages séparés pourra être due au hasard. D'ailleurs la notion de nauges séparés n'est pas aussi claire qu'il paraît d'abord. En résumé, même au seul point de vue théorique, la notion de type ou d'espèce est devenue, grâis à ce qui précède, plus précise, sans être encore complètement élucidée.

RELATIONSHIPS AMONG CHARACTERS OBSERVED IN UREA CLEARANCE TESTS

H. FAIRFIELD SMITH

Institute of Statistics
North Carolina State College

IN LABORATORIES of clinical pathology the observation usually made to measure efficiency of kidney action is that known as the van Slyke urea clearance test. Failure of renal function results eventually in increased concentration of urea in the blood; the objective of the test is to provide a measure of partial failure before the onset of that consequence.

It has been supposed that this can be obtained by observing the rate of urea excretion in urine. Following observations (under somewhat limited and special conditions) that urea excretion is approximately proportional to concentration in the blood, the measure adopted for test purposes has been the ratio of urea excreted per minute/urea concentration in blood—usually described as measuring the volume of blood "cleared" of urea per minute, although of course no blood is actually cleared.

A second consideration which the van Slyke test was devised to meet is that rate of urea excretion depends also on the rate of water excretion, particularly when this is slow. It was observed that rate of urea excretion is maintained nearly constant when rate of urine excretion exceeds about 2 ml. per minute. Consequently "maximum urea clearance" was defined as

$$C_m = UV/B \qquad (V > 2) \tag{1}$$

where V = volume of urine excreted in ml. per minute
U = concentration of urea in urine measured as mgm. N per ml.
B = concentration of urea in blood measured as mgm. N per ml.

In their early work van Slyke and his colleagues observed that at lower rates of urine excretion the quantity of urine urea (UV) varied approximately in proportion to $\sqrt{V}$. Using this relationship to estimate quantity of urea excreted when $V = 1$, they defined "standard urea clearance",

$$C_s = U\sqrt{V}/B \qquad (V < 2) \tag{2}$$

as an estimate of the ratio UV/B at a urine excretion rate of one ml. per minute.

Text books of clinical methods usually quote formulae (1) and (2), with the accepted criteria for each, but with no indication of their derivation. This was the only information available when in a prisoner of war camp in Siam an attempt was made to run some such tests as described by Wilshaw, 1946. The men observed were not expected to be renally defective, having been sent for test only as a precaution prior to operation. At first sight the results appeared alarmingly low, but it soon became apparent that they could not be interpreted according to the usual criteria. Therefore after checking chemical procedure, a group of unquestionably normal men were observed with intent to develope our own standards. Having no clue to the reason for the two usual formulae it seemed advisable to derive also our own adjustment for rate of urine excretion. For this purpose a linear relationship between log U and log V was found satisfactory.

When access to literature was again possible it seemed of interest to see if other published data might indicate that the method of adjustment used in Siam, or a similar one, could be generally adopted to provide a single criterion for any rate of urine excretion in place of rather artificial partition of observations into two groups with different criteria. Using the regression given below that can be done; but in the course of the work two fallacies underlying the test became apparent.

1. The correlation of urine and blood ureas which provides the rationale for the test was observed only for response *within* individuals to temporary experimentally imposed alterations of blood urea concentration. Observations on prisoners of war on deficient protein data, corroborated by other observations in India, showed that as *between* individuals on permanently different diets this correlation is not operative. Consequently the comparison to blood urea does not make allowance for differences in diet, as would often be necessary for validity of the test as usually applied.
2. The supposedly established criteria for normality were based on inadequate sampling procedure. They were merely means and ranges observed for a haphazardly selected group of 19 individuals (presumably workers around the laboratory). Such a sample gives little information on variation in the population at large. Furthermore, patients suspected of renal deficiency are commonly put on low protein diet and of such these nineteen are no sample at all.

These aspects have been discussed in a separate paper (Smith, 1949). Since the basis of the test is thus open to serious question the original objective of the statistical work here reported is irrelevant for immediate application. However, in so far that a statistical summary of existing data may aid further investigation, it seems advisable to make

available a brief account of the work which has been done. Should it still be considered desirable, for any purpose, to make comparisons of urine urea concentrations (U) or urea excretions (UV), either for the same person at different times or between persons, a method is suggested by which observations may be adjusted for rate of urine excretion, for alterations of blood urea due to *temporary* fluctuations (but not for permanent differences), and for size of person (required only when persons of appreciably different size are to be compared). But it is suggested that such comparisons be not considered as tests for renal defect, at least until the effects of diet have been more thoroughly elucidated.

INTERNAL REGRESSIONS

The first objective was to establish the relation between urea concentration and volume of urine. Figure 1 shows a typical set of data for one individual at approximately constant blood urea concentration. The procedure of van Slyke and his co-workers was equivalent to approximating the relationship by two straight lines

$$\log (U/B) = \log C_s - \tfrac{1}{2} \log V \qquad (V < 2) \tag{3}$$

$$\log (U/B) = \log C_m - \log V \qquad (V > 2) \tag{4}$$

Their point of intersection has been called the "augmentation limit", and has been discussed as a feature of potential interest to characterise different persons. As shown by Figure 1 these lines are not unreasonable as a first approximation, and (4) may be regarded as an asymptote for high V. It may be assumed, however, that the true curve should have no discontinuity; nor can (3) be regarded as asymptotic to the other extreme because all adults on whom more than ten observations have been made show that slope of log U on log V becomes numerically less than 0.5 when V is less than 1 ml. per minute (as has been noted also by Chesley, 1937, 1938, and by van Slyke, 1947). Consequently the "augmentation limit" is an artifact which cannot be regarded as a suitable statistic by which to describe curves exhibited by different persons.

The true form of relationship being unknown*, available data can be

*A curve asymptotic to (4) would appear to have some theoretical advantage. Several such were tried of which the best appeared to be (figure 1)

$$\log U = A - \log (a + V) + k \log B. \tag{5}$$

However, in the absence of knowledge of the true relationship, such forms appeared to give little advantage over (5), which is simpler for statistical purposes although unrealistic for large values of V beyond the point where $du/dv = m + 2cv = -1$. Within observable range such defect is trivial (figure 1).

FIG. 1

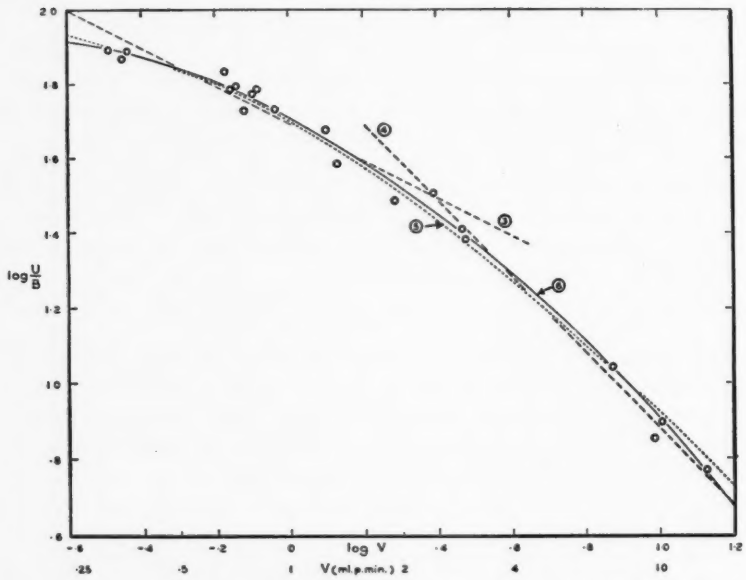


FIG. 1—OBSERVATIONS ON JHA, DATA FROM AUSTIN *ET AL* (1921)
Curve (3): log (U/B) = log 49.8 − ½ log V
Curve (4): log (U/B) = log 77.2 − log V
Curve (5): log (U/B) = 1.957 − log (.80 + V)
Curve (6): log (U/B) = 1.710−.5191 log V − .2865 (log V)²
(3) and (4) represent curves fitted by van Slyke *et al*, and correspond to C_s = 49.8, C_m = 77.2, respectively. This subject showed $k \simeq 1$.

Variances of deviations of log U from the regressions are:
for (3) and (4) .00127, for (5) .00113, for (6) .00104.

empirically described over the observable range and within limits of random variation by a polynomial of form.

$$u = a + mv + cv^2 + kb \tag{6}$$

where, for brevity, we write $v = \log V$, $u = \log U$, $b = \log B$. Common logarithms have been used.

All three variables, U, V and B being observed "as they come", in trying to evaluate the relationship between them there arises the question, as posed by Winsor (1946): "Which regression?" Throughout this work I take the point of view that our purpose is to examine what values of U occur with given observations of V and B (*cf.* postscript). Consequently I adopt the procedure of treating U as the dependent

variate and evaluate its regression on V and B as independent variates.** Although no definite evidence is available, it seems that random variation may be roughly proportional to magnitude of observation; therefore log observations are taken to have equal weight.

To seek evidence for the possibility that regression on blood urea might vary with rate of excretion (or conversely), the equation

$$u = a + mv + cv^2 + kb + g(bv)$$

was fitted to data for four men and for dogs reported by Austin *et al* (1921). No evidence for that possibility was detectable.

A separate regression of form (6) was fitted for each of the ten subjects for whom Austin *et al* (1921) and Möller *et al* (1928) report more than ten observations. (Actually two sets of observations, treated here as if for separate subjects, were on the same person in two widely separated periods, 1921 and 1928.) Nearly all drank varying amounts of water to produce variation in rate of urine excretion. Five (1928) ate urea to increase blood urea content on some days. There was some indication that, apart from effects of urea consumption, blood urea concentration may have tended to decrease with advancing time of day and when much water was drunk; that it varied more from day to day than within days, and was nearly as variable for one person between days as between persons. It would be of interest to discriminate between normal variation and superimposed effects of each factor; but as the experiments were not planned so as to facilitate such a breakdown it is scarcely practical to attempt it.

As a preliminary to further analysis we want to know whether or not all sets of observations show similar random variability. For variances of deviations about each regression, Bartlett's test (with notation of Thompson and Merrington, 1946) gives:

for all ten variances $M = 27.85$, $P < .01$,

excluding 2 subjects (Add and McL), $k = 8$, $M = 12.22$, P slightly $> .05$

excluding 3 subjects (Add, McL and WN) $k = 7$, $M = 6.18$.

The data for Add and McL had been previously reported from other

**It seems not to be usually recognized, and should therefore be pointed out, that even if we ascribe errors to V and B and fit a curve by the procedure described by Edwards Deming (1943), ignoring second and higher powers of residuals, we still arrive at the same curve as with the simpler assumptions.

laboratories and may therefore have been subject to different experimental variability. There is less justification for separating out WN, but some other features of his data also appear suspicious. If the statistical analysis were done on the spot one would like to repeat observations on this subject to see if he would persist in slightly exceptional behaviour. In the following, data for Add have been weighted by $\frac{1}{3}$, those for McL and WN by $\frac{1}{2}$, these weights being, in round figures, approximately proportional to reciprocals of the ratios of the individual variances to the mean variance of the other seven.

TABLE 1

Coefficient	Range	Weighted mean*	Variance ratio, F. ($n_1 = 9$, $n_2 = 127$)	P
u' (at $v = .3$, $b = -.7$)	.70 to .94	$.815 \pm .022$	7.00	$< .001$
m' (at $v = .3$)	$-.79$ to $-.60$	$-.701 \pm .019$	2.28	$ca.025$
c	$-.62$ to $-.16$	$-.355 \pm .050$	3.87	$< .001$
k	.27 to 1.14	$.802 \pm .081$	2.98	$< .005$

*Standard errors determined from mean squares *between* persons.

Figure 2 shows the fitted regressions on v, and table 1 summarises variation of the regression coefficients as computed for the form

$$u = u' + m'(v - .3) + c(v - .3)^2 + k(b + .7) \qquad (7)$$

That is, we compare urine urea concentrations (u'), and their rates of change with v, (m'), near the middle of the range of observations, namely at $v = .3$, $b = -.7$. (Over-all averages are $\bar{v} = .313$, $\bar{b} = -.755$. If u' and m' were evaluated for a non-central origin as in (6) they would be subject to higher errors. Coefficients c and k are not affected by change of origin for either variate, m' is affected only for v.) The last two columns show tests of significance for heterogeneity of each type of coefficient after fitting 30 others (each of the other three for every individual). The mean squares for each coefficient are not independent; a test for heterogeneity of m, c, k collectively is given later (table 3). Curves for the subject who was observed during two periods (1921 and 1928) are marked A and B in Figure 2. Both show less curvature than all others ($c = -.16$ and $-.20$ as compared to $-.29$ to $-.62$ for others), indicating that such characteristics may persist over long periods.

FIG. 2

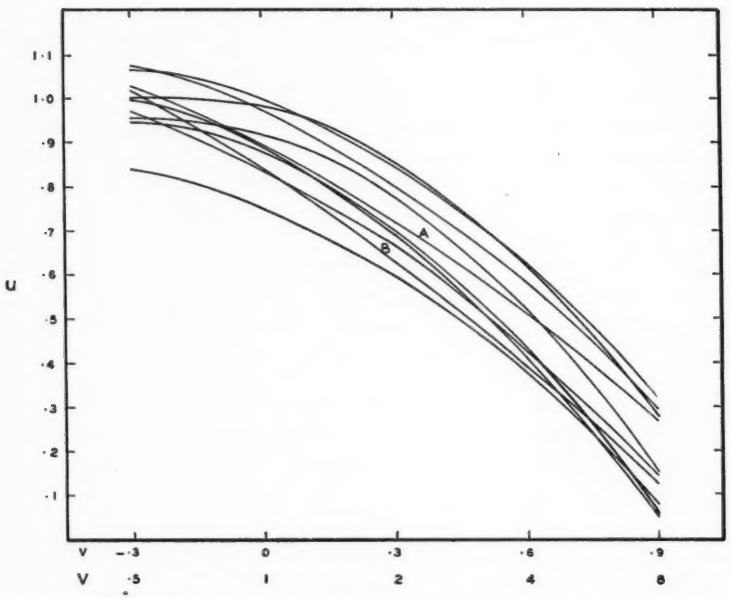


FIG. 2—REGRESSIONS ESTIMATED FOR 10 ADULTS INDIVIDUALLY. RANGES AND MEANS OF THE REGRESSION COEFFICIENTS ARE INDICATED IN TABLE 1. VERTICAL POSITIONS HAVE BEEN ADJUSTED TO BLOOD UREA = .15 MGM. *N* PER ML., HEIGHT = 170 CMS. DATA FROM VAN SLYKE *ET AL* (1921, 1928).

Möller *et al* report 2, 3 or (for one subject) 5 observations on each of nine other subjects. For prisoners of war (15 "normals", 50 "patients", classified as indicated above) we have two observations of U and V in successive hour periods, but only one observation of B which can be assumed constant for both hours. No internal regression on b can therefore be evaluated for the POW; and for neither group can we fit a regression with respectively four and three coefficients for each individual. But we can get an average regression, as in ordinary applications of analysis of covariance, from pooled sums of squares and products within subjects. Deviations from the average curve will, in these cases represent variation between persons. Even when there are more than four observations per person it would seem preferable to evaluate the average regression for a group in this way rather than by averaging individual coefficients; because when observations of independent variates are differently distributed there exist variable correlations between

the coefficients for each, and because slopes on v in the varying ranges contribute information on the quadratic coefficient (the only source of an estimate of curvature for an intraclass regression with just two observations per class).

Average regressions thus determined are presented in table 2 for five groups of adults, *viz.* observations made in each of three American laboratories and the two groups of POW. Variation is summarised in table 3. Mean squares for variation of regressions among individuals within each of three groups (subdivision of row 3) are not detectably heterogeneous; subdivision of row 4 is discussed above. Row 2 compared to row 3 indicates heterogeneity of average regressions of the five groups, significant at the 5 percent level. However, considering the extreme differences of environment, with POW subject to extreme protein deficiency, the very small degree of heterogeneity is remarkable; furthermore most of it is ascribable to the single individual observed by Addis and not to contrast between Americans and POW, (table 2). (The latter contrast does not involve regression on b.) Row 3 *versus* row 4 indicates significant heterogeneity of regressions for individuals

TABLE 2

Data	No. of Individuals	No. of Observations	m at $v = 0$	m' at $v = .3$	c	k
1. Addis	1	29	−.401	−.743	−.570	.649
2. McLean	1	17	−.445	−.650	−.341	.952
3. Van Slyke *et al*: adults	17	144	−.469	−.694	−.375	.847
4. POW: "normals"	15	30	−.555	−.646	−.152	—
5. POW: "patients"	50	100	−.488	−.640	−.253	—
6. Average all adults above	84	320	−.489	−.695	−.344	.869
7. McIntosh *et al*: children	8	48	−.785	−.783*	−.017	.299
8. Austin *et al*: dogs	2	26	−.463	−.826	−.605	.746
Standard errors: Row 1			.059	.032	.073	.146
2			.088	.129	.279	.130
3			.021	.011	.024	.039
4			.071	.051	.097	—
5			.066	.047	.088	—
6			.020	.0114	.024	.044
7			.027	.026	.053	.190
8			.108	.105	.217	.091

*Linear regression on v, the quadratic term being negligible.

TABLE 3

	d.f.	M. Sq.
1. Mean regression (within all adults)	3	6.44363
2. Heterogeneity of regressions of 5 groups (table 2)	10	.01181
3. Heterogeneity of regressions among persons within groups	96	.005050
4. Deviations from regressions within persons	127	.002057
Sub-division of row 3:		
van Slyke *et al*: 17 persons	35*	.003650
POW normals 15 persons	13	.002626
POW patients 50 persons	48	.006724
Sub-division of row 4:		
van Slyke *et al*: 8 persons	89	.001953
Addis: 1 person (weighted by 1/3)	25	.002333
McLean: 1 person (weighted by 1/2)	13	.002234

*Contains 1 d.f. which properly belongs in row 4, being a deviation from regression for the subject with 5 observations, whose regression was not separately evaluated.

observed in the same laboratory ($P < .001$), but again the actual magnitude of differences is small.

Children observed by McIntosh *et al* (1928) show similar variation (table 5). Their average regression on v appears approximately linear with slope corresponding to that for adults at about $V = 2.6$ ml. per minute. This suggests that there should be progressive alteration of m with increasing age or size. No such correlation could be demonstrated; but, with only 5 to 10 observations per child, individual coefficients were erratic. Their regression with blood urea is low: they were not given supplementary meat or urea.

Dogs (Austin *et al*, 1921) show a similar regression with no coefficient significantly different from those for adults. The experiment on dogs was run to test correlation of urine urea with blood urea as temporarily varied by consumption of urea or meat. For both dogs and men this regression is significantly less than unity, the value which has been postulated by previous workers. (Table 1 shows that the mean discrepancy appears still significant when tested against an error based on variation between individuals.)

If it be desired to estimate from an observation of urine urea concentration what it would have been at some other rate of urine excretion, say V_0, most of any required adjustment will be determined by the linear coefficient on v. Although regressions exhibited by different persons vary significantly, table 1 suggests that the linear coefficient

for variation around average v is moderately consistent. Consequently, for any observed V not widely divergent from V_0 , adjustments appropriate for different persons should be not dissimilar, and may, for most purposes, be determined from a well established average curve. If there are only one or two observations for a given person, this will be the only method of evaluation available. Even with more observations, unless unusually numerous, evaluation of an *ad hoc* regression would be unduly laborious and often less accurate to boot (particularly if V_0 should lie outside the range of observations). Regressions on blood urea appear somewhat more variable; but their average appears stable and normal variation of B is relatively small, therefore similar arguments may apply. We therefore give in table 2 the average regression for all adults. Being derived from observations on both Americans and POW'S this regression would appear to have a wide inductive basis. Nevertheless it should be noted that neither group is a satisfactory sample of a normal population. Therefore although the regression may be considered well established and may serve immediate purposes, if extensive use is to be made of it it should be re-evaluated for some sample properly representative of a population to which it is to be applied.

Using the regression coefficients of row 6 of table 2, urine urea concentrations at a fixed rate of urine excretion (V_0) and at blood urea concentration (B_0) may be estimated from

$$\log \hat{U} = \log U + .49(\log V - \log V_0) + .34\{(\log V)^2 - (\log V_0)^2\} - .87(\log B - \log B_0)$$

Where U, V, B are observed values. If the adjustment for B is included it should be noted that this will estimate urine urea only for B changed to B_0 by a *temporary* alteration in diet; the regression does not hold for long term alterations. If it is required to adjust a mean of several observations, means of logarithms should be used and the value inserted for $(\log V)^2$ should be the mean of $(\log V)^2$ and not the (mean of $\log V)^2$.

If comparison is to be made between persons of different sizes a further term, $-1.5\ (\log H - \log H_0)$, may be added, where H = height (see below).

EXTERNAL REGRESSIONS

By a sub-division of analyses of variance, similar to that given by Kendall (1948, Sec. 24.30), we may examine whether variation between men indicates any different relationship to the independent variates as compared to that within. This analysis is given in table 4. For van

Slyke's adults we consider heterogeneity of regression on the three variates (v, v^2 and b); POW's, for the reason given above, show internal regression only on v and v^2. Besides these table 4 shows the part of sums of squares between men attributable to regression on an extra variate for which there are only single observations per man; for van Slyke's adults on height, for POW's on blood urea. (Heights of POW's were not recorded.)

TABLE 4

	v. Slyke *et al*: adults		Prisoners of war normals		patients	
	d.f.	M. Sq.	d.f.	M. Sq.	d.f.	M. Sq.
Heterogeneity of regressions "within" and "between"	3	.02007	2	.00095	2	.04295
Subsidiary regression (see text)	1	.02408	1	.04413	1	.04310
Deviations from regression between men . .	12	.02221	11	.01817	46	.04280
Total between men ("reduced" for adjustment by internal regression)	16	.02193	14	.01756	49	.04281
Deviations from average internal regression .	124	.00243	13	.00263	48	.00672

No different relationship between as compared to within men can be detected. The enhanced mean square for adjusted means apparently represents individual variation occurring at random relative to the independent variates investigated.

Between POW's no correlation with blood urea concentration can be detected. (The condition that the mean square for this regression in the group of normals is greater than the remainder mean square—albeit not significantly so—is due entirely to a single observation. For both groups together the correlation would be approximately zero. See figure elsewhere—Smith, 1949.) For van Slyke's adults this regression is similar both between men ($k = .837 \pm .109$) and within ($.847 \pm .039$). Since reasons have been given in the earlier paper for believing that the normal relationship between individuals on unequal stable diets may be as in the POW's, the relationship shown by van Slyke's data calls for comment. Major differences between the blood ureas of his subjects

were due to some having consumed urea, and consequently reflect the same temporary fluctuations which determine the internal regressions. Among means of thirteen persons who were not given urea the regression, after adjustment for V, is $k = .38 \pm .22$. It seems likely that persons on normal diet may show some, but only low, correlation between blood and urine urea concentrations. If it be still considered worth while to compare urine urea concentrations this point requires further investigation on a larger sample which is properly representative of a definable population.

CORRELATION OF UREA EXCRETION WITH BODY SIZE

For most adults the effect of body size on urine urea concentration at fixed V is usually negligible relative to random variation. But for persons divergent from normal (taken as 170 cm. in height) adjustment for size has been recommended. Möller *et al* (1928) presented reasons for considering that the best characteristic for this purpose is surface area, which they computed from height and weight by the formula of Du Bois. Later McIntosh *et al* (1928) argued that weights of patients suffering from disease are unreliable as an indication of their normal sizes, and they prefer to base size adjustments on height alone. Their procedure is, however, circuitous and confusing.

From height they estimate what they term the body weight "ideal" for height, by which, apparently, they mean average body weight for a population of given height. The estimate of average body weight thus obtained is then recombined with the height measurement to estimate surface area. Actually the only variate is height, and the net effect is the same as if adjustment were based directly on it alone. To introduce a supposititious surface area is merely to mask what is in fact the basis for the corrections, and to complicate computations.

To investigate whether weight, either in its own right or as an indicator of surface area, may add to the accuracy of adjustment, more information may be gained by considering both variates (height and weight) explicitly, rather than arbitrarily combined in a formula under another name. The experiments on children (McIntosh *et al*, 1928) were planned specifically to consider effect of body size. After adjusting means per child by their internal regression on v and b, regression on size characters gives analysis of variance as in table 5. This shows that once adjustment has been made for height, nothing further is gained by considering weight. (Correlation with log weight alone is nearly, but not quite, as great as with log height. It would account for .708 in the sum of squares, with regression coefficient $0.97 \pm .24$.)

Variation in size of the 17 adults observed by van Slyke *et al* is too

TABLE 5

	d.f.	S. Sq.	M. Sq.
Between children (adjusted for v and b) . . .	7	.9691	
Regression on log height	1	.7881	
Deviations from regression on height . . .	6	.1810	.03018
Partial regression on log weight	1	.0107	
Within children			
Heterogeneity of internal regressions . . .	21		.00712
Deviations from individual regressions . .	23		.00274

small to show significant correlation (subsidiary regression, table 4). However, the regression is not dissimilar to that observed for children, and here also correlation with height was slightly greater than with weight.

Regression coefficients of u on log height are:

for children	$1.66 \pm .33$
for adults	$.96 \pm .92$
weighted mean	1.59

Since surface area as computed by van Slyke *et al* appears to be proportional to $(\text{height})^{1.40}$ for children, and to $(\text{height})^{1.46}$ for adults the hypothesis on which their adjustment is based appears reasonable.

POSTSCRIPT

A referee, on reading an earlier version of this paper, suggested alternative formulation which appears of interest and worth comment. He suggests that the important quantities to be considered are:

total urea excreted per minute $= UV$
concentration potential $= \log U - \log B$,
volume rate $= V$,
and normal liquid handling capacity of the kidney tubules:

and he asks, very pertinently, why was U chosen as the significant dependent variate?

Relative to the van Slyke test the total amount, UV, is more important than the concentration U. But V being still the independent variate, assumed for purposes of curve fitting free of error, to use either u or $(u + v)$ as dependent variate leads to identical regressions on v

(except for a difference of unity in the linear coefficient). Consequently it is simplest to deal with the variates observed. According to convenience in any particular context the regression may still be expressed in either form.

Secondly he suggests that if ratio of volume excreted to normal capacity may be important, and if we may assume normal capacity $\propto$ weight $\propto$ height^3, we should use as independent variate V/H^3. Elsewhere (1949) I have criticised the *a priori* assumption that the ratio U/B is significant, and (1936) pointed out that to use arbitrary ratios may usually be inefficient. The first need it to establish the relative effects of both variates. If confidence intervals of regression coefficients include values conforming to some combination of the variates which has theoretical backing, then the combination may be adopted if it gives a simplification. For investigation one could use (V/H^3) and either V or H as the two variates. But it seems simpler initially to use the primary variates actually observed, rather than to focus attention on a particular one of many possible hypotheses. We may then consider what would be the effect on the straight-forward regressions if some compound of the variates were in order. Linear combinations present no problem since corresponding terms of the regression can be freely re-arranged as in the preceding paragraph. To examine the effect of introducing a quadratic term such as $c(v - ah)^2$—where $h = \log H$, a is some constant and we suppose $c \neq 0$—, we would note that, if this may be a better fitting regression than (6), then m (or m') for each individual should show regression on h with slope $-2ac$. Secondly the partial linear regression on h when evaluated in the usual way should be $-am'$, where $m' = du/dv$ near the mean of v, for our data about $-.7$. Therefore, if (V/H^a) is to be used as independent variate, both adults and children indicate that a should be about 2. However variation of H among adults is too small to show any effect of incorporating H in the quadratic term, and for children c is approximately zero so that the whole term drops out anyway. Therefore no evidence as to the merits of this hypothesis can be deduced from present data.

SUMMARY

Urea clearance tests based on the van Slyke formulae were found to be very misleading when used in a prisoner of war camp where diet was deficient in protein. Subsequent comparison of the POW observations with others recorded in the literature revealed that the test appears to have been based on (statistically) false premises. Before this was realized considerable statistical analyses had been done, both of the POW data and of those on which the test was founded, with a view to

simplify the test. Although that purpose is now irrelevant an account of the work is given here in the hope that it may aid further research.

The relation between urea concentration in urine and volume of urine excreted in a given time varies significantly between one person and another. All, however, show similar curves which can be approximated by a quadratic in logarithms of the variates.

The regression of log urine urea on log blood urea concentration is about .85 for temporary fluctuations. This is significantly less than unity as postulated by previous workers. Between persons on permanently different diets these variates may be uncorrelated.

Urine urea concentrations of children at fixed urine excretion rate vary approximately as $(\text{height})^{3/2}$, or in proportion to surface area as postulated by previous workers.

ACKNOWLEDGEMENTS

The observations on prisoners of war were made by Major A. T. H. Marsden and Major R. G. H. Wilshaw. Mr. Fan Kam Yin, of the Rubber Research Institute of Malaya, assisted with the computations. I am indebted also to Professor John W. Tukey for stimulating discussion.

LITERATURE CITED

Austin, J. H., Stillman, E. and van Slyke, D. D. Factors governing the excretion rate of urea. *J. Biol. Chem.* 46: 91, 1921.

Chesley, L. C. The validity of the calculation of standard urea clearance from low urine volumes. *J. Clin. Invest.* 16: 625, 1937.

——— Urea excretion at low urine volume. *J. Clin. Invest.* 17: 119, 1938.

Deming, W. Edwards. *Statistical adjustment of data.* Wiley & Sons, New York, 1943.

Kendall, M. G. *Advanced theory of statistics.* Vol. 2. Griffin, London, 1948.

McIntosh, J. F., Möller, E. and van Slyke, D. D. Studies of urea excretion III. The influence of body size on urea output. *J. Clin. Invest.* 6: 467, 1928.

Möller, E., McIntosh, J. F. and van Slyke, D. D. Studies in urea excretion II. Relationship between urine volume and the rate of urea excretion by normal adults. *J. Clin. Invest.* 6: 427, 1928.

Smith, H. Fairfield. Urea clearance tests. *J. Clin. Path.* 2: 266, 1949.

——— Discriminant function for plant selection. *Ann. Eug.* 7: 240, 1936.

Thompson, C. M. and Merrington, M. Tables for testing the homogeneity of a set of estimated variances. *Biometrika* 33: 296, 1946.

Van Slyke, D. D. The effect of urine volume on urea excretion. *J. Clin. Invest.* 26: 1159, 1947.

Wilshaw, R. G. H. Improvised biochemical work of a prisoner of war. *The Pharmaceutical J.* 24 Aug. 1946. Reprinted: *Malayan Agric. J.* 30: 27, 1947.

Winsor, C. P. Which regression? *Biom. Bul.* 2: 101, 1946.

ON THE ERRORS OF BIOLOGICAL ASSAYS WITH GRADED RESPONSES, AND THEIR GRAPHICAL DERIVATION*

NORMAN T. GRIDGEMAN

Food Research Department
Lever Brothers & Unilever Limited
Colworth House
Sharnbrook, Bedford, England

INTRODUCTION

CALCULATION of the fiducial limits of error of a bioassay is intricate. A technical assistant, computing by rote, is unlikely to be able to decide whether an abnormal result is due to miscalculation or is inherent in the data. Nomography and the use of error curves can reduce his labors, promote his confidence, and advance his insight into the factors contributing to the assay error.

Several graphical aids for bioassays and errors are already available, e.g. those of de Beer [1], Knudsen & Randall [2], Bliss [3], Sherwood [4], Goyan & Dufrenoy [5], Koch [6], and Harte [7]. Lees's specific calculating machine [8] should also be mentioned in this connexion. The most recent aid, and perhaps the one most nearly suited to the requirements of the writer's laboratory, is that of Healy [9], whose nomogram gives errors for assays of given size and structure. The present requirements, however, were error charts covering 4- and 6-point assays of given design but greatly varying size. It was not required to calculate graphically the mean estimate itself, the writer being unable to appreciate the advantages of so replacing the simple arithmetic operations involved.

The devisal of appropriate error charts, the main subject of this paper, has thrown into relief some aspects of assay design and error factors that may be worthy of concomitant record.

ERROR THEORY

To explore the graphical possibilities of the error distribution, let us derive a comparatively simple general expression for the fiducial

*Presented at a joint meeting of the Biometric Society (British Region) and the Biological Methods Group of the Society of Public Analysts, London, March 14, 1950.

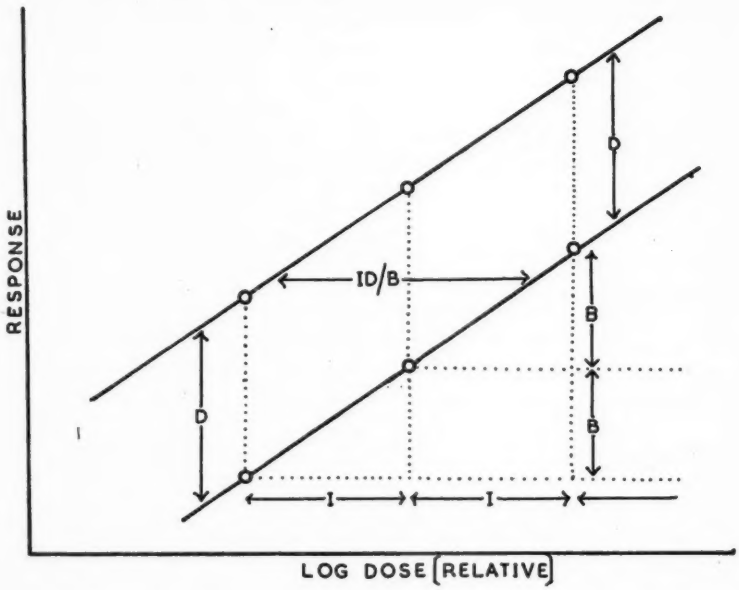


FIGURE 1.
Showing the geometric relations between I, D and B for a 4- or 6-point (or any $2k$-point) assay. Either set of linear points may represent the standard preparation (S), the other then representing the test material (U).

limits of a continuous response assay with equal numbers of responses in each of $2k$ dosage groups, where k is any integer other than unity. Half of the groups shall be on the standard preparation and half on the test material, and there must be a uniform potency ratio, antilog I, between the doses for adjacent groups of either preparation. Let n be the total number of responses and s the pooled estimate of the homogeneous standard deviation of responses within groups. Assume a linear relation between response and log dose. Let B be the mean response increment between adjacent dosage groups of either preparation (i.e. the coefficient of regression, or slope, b, is B/I), and let D be the corresponding mean increment of all "test" responses over all "standard" responses. The fraction D/B, which we shall make $= A$, determines the best estimate R of the unknown potency of the test material in terms of the standard (dose for dose); for $R =$ antilog $M =$ antilog IA. Figure 1 illustrates these relations.

Now A can be regarded as a sample of the population value α belonging to an infinitude of assays identical in size, design etc., with the

actual one. We can restate this as the long-run tendency for αB to be equal to D, and this consideration provides the clue, first followed up by Fieller [10], to the required fiducial limits of A and, consequently, of R. The difference $D - \alpha B$ is normally distributed about a mean of zero and therefore, when measured in units of its estimated standard deviation, will follow the t distribution with the number of degrees of freedom associated with s. The estimate of the standard deviation of $D - \alpha B$ is

$$\sqrt{V(D) - 2\alpha Cov(DB) + \alpha^2 V(B)},$$

so that we may write

$$t^2 = (D - \alpha B)^2/[V(D) - 2\alpha Cov(DB) + \alpha^2 V(B)],$$

which is a quadratic in α:

$$\alpha^2[B^2 - t^2 V(B)] - 2\alpha[BD - t^2\, Cov(BD)] + D^2 - t^2 V(D) = 0.$$

This equation can be simplified if we introduce Fieller's statistic,

$$C = B^2/[B^2 - t^2 V(B)],$$

and make use of the facts that, in assays of the type we are here concerned with,

$$V(D) = 4s^2/n, \qquad V(B) = 12s^2/n(k^2 - 1),$$

and the covariance of B and D nil. It follows that

$$B^2 - t^2 V(B) = B^2/C,$$

and that

$$t^2 V(D) = B^2(C - 1)(k^2 - 1)/3C.$$

Substituting these in the above equation, multiplying throughout by C/B^2, and replacing D/B by A, we finally arrive at

$$\alpha^2 - 2\alpha CA + CA^2 - (C - 1)(k^2 - 1)/3 = 0,$$

whose roots are

$$CA \pm \sqrt{[C - 1][CA^2 + (k^2 - 1)/3]}.$$

The roots give the upper and lower fiducial limits, i.e. the range whose probability of not enclosing α is P, a value determined in the choice of t and usually taken at 0.05. To obtain the percentage fiducial limits we must use

$$A(C - 1) \pm \sqrt{(C - 1)[CA^2 + (k^2 - 1)/3]}, \qquad (1)$$

which gives two values that, multiplied by I and added to 2, will be the logs of the required percentages.

Several points of interest attached to this expression are worth special mention:

(i) When A is zero, i.e. when the mean responses to the standard preparation and the test material are equal, it reduces to

$$\pm \sqrt{(C-1)(k^2-1)/3},$$

and illustrates the biometrical advantage of shrewd guessing at the potency of the test material, and consequently of so choosing the dosages to bring A as close to zero as possible. It is also apparent that only when A is zero will the fiducial limits of the logarithm of the assay be symmetrical about the mean; as A moves away from zero a progressive bias sets in. Note that the expression takes the very simple form $\pm \sqrt{(C-1)}$ for a 4-point assay with equal responses.

Theoretically A has no upper or lower limit, but in practice it is seldom that A can exceed ± 2 without straining the validity of the assay, usually by a breakdown of the parallelism of the "standard" and "test" response curves.

(ii) Because in theory the variance of the slope of the mean response curve can have any value including 0, as $t^2V(B)/B^2$ passes from zero in a positive direction, so C passes from unity to $+\infty$ and then from $-\infty$ to zero, following the type curve $y = 1/(1-x)$. As C approaches unity the value of the expression dwindles towards vanishing point where the precision of the assay apotheosizes at 100%. A value of infinity, or even a negative value, for C is, by contrast, realizable. At infinity the position is interesting: it is reached when $B^2 = t^2V(B)$, i.e. when the slope of the mean response curve is exactly poised between significant and non-significant differences from zero at the selected probability level. It is then obvious that the assay is almost worthless, and this is reflected in the shift of the bias (the mid-point of the error range) to infinity, and the expansion of the range itself to nearly $\pm\infty$; in fact one of the two values of our expression becomes infinity and the other a comparatively small negative whose magnitude depends on A and k. This latter value can readily be found if we equate expression (1) to L and rearrange the equation so that C is collected. We have

$$1 - 1/C = L^2/[(L+A)^2 + (k^2-1)/3];$$

whence, if $C = \infty$, we obtain

$$L = -[A^2 + (k^2-1)/3]/2A$$

for the required value of L. Remembering that the other value is infinity (positive or negative according to the sign of A), and that we have to take antilogs in the transformation of L to limits of error, we can see that when A is positive (i.e. when the test material is more potent than hypothesized) the upper fiducial limit will be infinity but the lower limit will have a finite value. On the other hand, when A is negative (i.e. when the estimated potency is low) the lower limit will be nil but the upper limit finite. Only in the special case of $A = 0$ will the range extend from 0 to ∞. Reflection on this critical state of slope significance will show that the emergence of one finite limit is eminently reasonable. If, for example, a material expected to contain 100 units per gram assays at 200 units per gram, with a slope of enigmatic significance at $P = 0.05$, we cannot expect to extract any information on how *large* the true unitage may be, yet we can be fairly confident that the material has *some* potency, and we shall not be surprised to find that, in fact, a unitage is calculable below which there is only a 2.5% probability of encountering the true value.

Occasionally C may be negative; this means that $t^2V(B) > B^2$ and that there are no real fiducial limits. (The equations for L are soluble with negative C's under certain conditions, but the roots so obtained are imaginary.) The assay is consequently entirely valueless at the working probability level.

(iii) The two commonest values for k are 2 and 3, so we have:

$$\text{for the 4-point design: } A(C - 1) \pm \sqrt{(C - 1)(CA^2 + 1)};$$

$$\text{for the 6-point design: } A(C - 1) \pm \sqrt{(C - 1)(CA^2 + 8/3)}.$$

The comparatively simple relation of these two expressions masks complex and contingent relations between the errors of the two designs for the same number of subjects (animals). One simple deduction can however be made: for given values of n, s and b a 6-point design with interval I tends to biometrical identity (interactions or validity tests aside) with a 4-point design with interval $I\sqrt{8/3}$ for all values of A; the tendency is a function of t, whose degrees of freedom will differ slightly for the two designs, the difference becoming negligible as n increases. In general the 4-point assay is more or less efficient than the comparable 6-point assay if the interval is $>$ or $< I\sqrt{8/3}$. This aspect of design is enlarged on in the Appendix.

(iv) Finally, if, as above, we let L denote a root of the expression, it is manifest that error curves applicable to given values of L can be drawn with C and A, or with functions of C and A, as co-ordinates. This is the basis of the charts discussed hereunder.

BACKGROUND OF CHARTS

Of the two co-ordinates for the projected L curves, A is a simple key datum that is normally calculated *en route* to the evaluation of the assay itself. It may, in fact, be preferable to graduate the A scale as antilog IA, which is R, the best estimate of the potency of the test preparation as a multiple of the expected potency (the null hypothesis), and this practice has been adopted here. On the other hand C, which is $B^2/[B^2 - 12t^2s^2/n(k^2 - 1)]$, is complex, a function of five variables. However, one of these, t, is wholly determined by n and the structure of the assay, and another k, can be fixed; this leaves three variables: B, s and n. Now if we take the trouble to calculate an appropriate function of any one pair of these, we shall be left with two values to derive C from, and this lends itself to simple nomography. The best pair for choice is undoubtedly B and s, because they are components of the important statistic λ, whose evaluation is commendable *per se*. Lambda is the residual standard deviation of the response measured in slope units i.e.

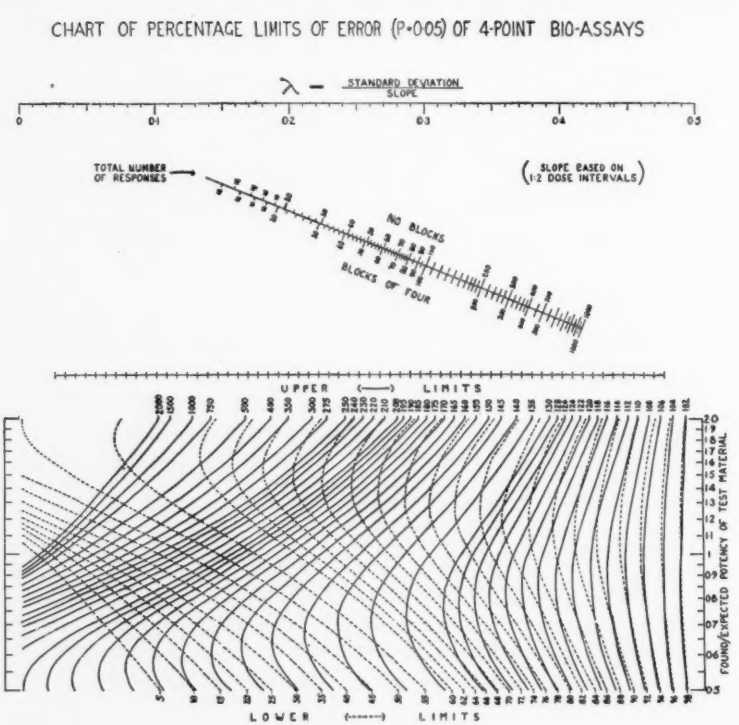
$$\lambda = s/b = s\,I/B,$$

and being a ratio, independent of the kind of response, of the size of the metameter, and of n (the size of the assay), λ is a coefficient of the inherent sensitivity of the particular set of animals used in an assay. To evaluate λ as a matter of routine, to keep track of its fluctuations over time, and to compare it with evaluations in other laboratories, is a useful discipline. To know the relative mean λ's of two types of assay conducted with the same kind of animal is to know the relative cost, as far as animal usage is concerned, of the two assays at any degree of precision. And the two components of λ are necessarily always available. Bliss and Cattell [11] have compiled a useful list of λ's for various kinds and specimens of bioassay.

Introducing λ, we may re-fashion the equation for C as,

$$C = 1/[1 - 12\lambda^2 t^2/nI^2(k^2 - 1)],$$

showing the basis of its derivation from two variables λ and t^2/n. In sum, then, for particular values of I and k, we can construct a nomogram to estimate C from λ and n, and use the C scale as an axis of the error curves. Thus for an assay of given design the percentage fiducial limits of error at a given probability level can be read off from the three data, λ, n and A. It may be well to emphasize that the chart as an entity, nomogram plus graphs, must have I and k determinate—we cannot, for instance, use one set of error curves in conjunction with two nomograms based on different I's and/or different k's.

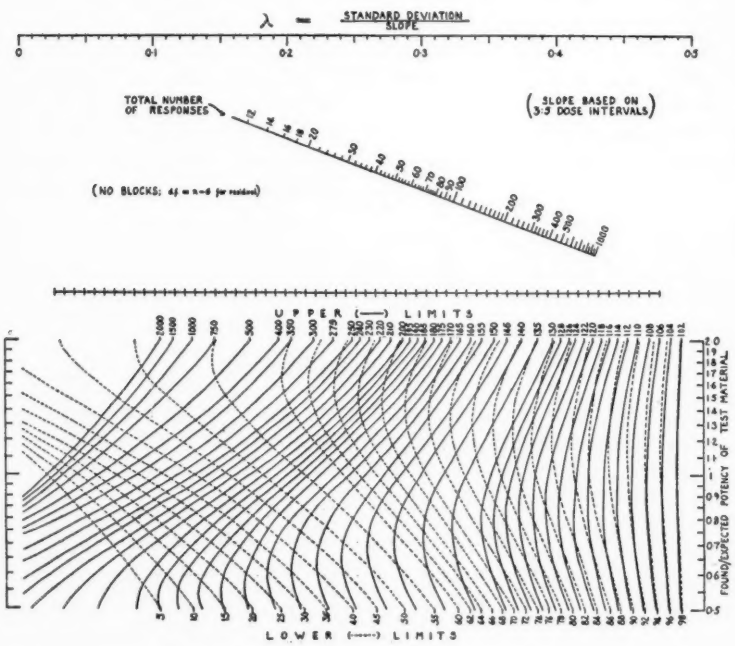


The two specimen charts herewith reproduced cover the 4-point assay with 1:2 "steps" ($k = 2$; $I = 0.30103$) and the 6-point assay with 3:5 "steps" ($k = 3$; $I = 0.22185$). The centre horizontal scales of $F(C)$ are graduated arbitrarily as in practice they act merely as gate-ways to the error curves. Choice of a suitable function of C turns on several considerations including the desirability of a linear distribution and the accommodation of suitable ranges and spacings of λ's, n's and the error curves. The best choice is $\sqrt{(C-1)/C}$, which is the square root of Finney's statistic g [10].

The nomogram for the 4-point assays has a double n scale: one for "no blocks", i.e. t is based on $n - 4$ degrees of freedom; the other for blocks of four (e.g. isogenic tetrads of subjects), t being based on $3n/4 - 3$ degrees of freedom. As blocks of 6 are rare in bioassay work, only the blockless design has been catered for on the 6-point chart.

Symmetry of design, i.e. equal numbers of responses at each point, is a theoretical condition of the use of either chart; nevertheless the

CHART OF PERCENTAGE LIMITS OF ERROR (P=0·05) OF 6-POINT BIO-ASSAYS



slight asymmetry brought about by a few lost responses will not seriously distort the error from that that would have been obtained by the same losses evenly distributed among the groups. Indeed, to judge from some examples given later in this paper, asymmetry must be considerable before the charts cannot safely be used to find the fiducial limits.

CONSTRUCTION OF CHARTS

The nomogram consists of two parallel uniform scales connected by a third intercept scale to form an N-shaped triad. If the parallel scales are zero'd at the junctions, the diagonal scale can be calibrated to yield ratios of aligned pairs of readings. For the present purpose it is meet to distribute λ on one parallel and $\sqrt{g}$ on the other, and from them to calibrate the diagonal in n (necessarily non-uniform). Then a straight edge laid across the diagram and connecting any λ with any n will cut the $\sqrt{g}$ scale at the correct value.

The size of the scaling must first be settled. Experience of the

order of magnitude of λ likely to be met with, together with considerations of the lay-out of the chart as a whole, indicate that a λ scale from 0 to 0.5, coupled with a $\sqrt{g}$ scale of 1 to 0, is suitable. It should be noted that $g = \sqrt{g} = 0$ corresponds to $C = 1$ and absolute precision, that $g = \sqrt{g} = 1$ corresponds to $C = \infty$ and only one real fiducial limit, and that $g > 1$ results in a negative value for C and absolute imprecision. A perpendicular drawn between these two scales will cut λ and $\sqrt{g}$ where the latter is numerically equal to $1 - 2\lambda$. Now as

$$\sqrt{g}/\lambda = t/I\sqrt{n(k^2 - 1)/12},$$

the point on the diagonal cut by the perpendicular must correspond to a unique value of n for a given I and k. This provides the key to the calibration of the diagonal. Replacing $\sqrt{g}$ by $1 - 2\lambda$ in the above expression, and re-arranging, we find

$$\lambda = I\sqrt{n(k^2 - 1)/12}/[t + 2I\sqrt{n(k^2 - 1)/12}],$$

from which any required n can be located on the diagonal as the intercept of the perpendicular from the corresponding λ. To locate, for instance, n's in the "no block" scale for $I = \log 2$ and $k = 2$ at the 5% probability level we substitute in

$$\lambda = 0.30103\sqrt{n}/(2t + 0.60206\sqrt{n}),$$

the t-table being entered at $n - 4$. Thus for $n = 20$ the value of λ is 0.1942; so we locate this n as the intercept from the perpendicular at this λ. In practice it is necessary to locate only a few n's by calculation and construction, the others being found by graphical interpolation.

The $\sqrt{g}$ scale must now be considered as one of the ordinates for the error curves to be drawn alongside. As the $\sqrt{g}$ scale thus acts merely as a stepping stone from the nomogram to the curves, in the finished diagram it need be neither labelled nor figured. The other ordinate shown on the finished charts is R, which is "found potency of test material as a multiple of the hypothesized potency", i.e. antilog IA, and the range chosen is 0.5 to 2, which will cover the great majority of assays. (In practice assays yielding values outside this range are generally repeated with an amended hypothesized potency.) The working scale however is the more amenable A, distributed linearly between ± 1 if $I = \log 2$, and between ± 1.357 if $I = \log 5/3$.

We require to plot A against $\sqrt{g}$ for given values of

$$A(C - 1) \pm \sqrt{(C - 1)[CA^2 + (k^2 - 1)/3]}$$

— or, more particularly, for given percentage limits of error at whatever probability level has been built into that expression and which

here happens to be $P = 0.05$. Let L be a root of the expression. Recalling that $C = 1/(1 - g)$ we can easily derive that

$$\sqrt{g} = \pm L/\sqrt{(L + A)^2 + (k^2 - 1)/3}.$$

Let us consider the curve of all combinations of A and $\sqrt{g}$ that yield an upper limit of 160%, i.e. 1.6 times the mean assay estimate, for 4-point designs ($k = 2$) and $I = \log 2$. The logarithm of 1.6, which is 0.2041, is IL. Therefore $L = 0.6780$. We can now calculate $\sqrt{g}$ for any value of A. For example:

A	$\sqrt{g}$
1.0	0.3471
0.5	0.4388
0	0.5612
−0.5	0.6675
−1.0	0.6454

The whole curve can be drawn from these five points. It is useful however to include a sixth, viz, the maximum, which always occurs at $L = -A$, and which for this curve is at $A = -0.678$ and $\sqrt{g} = 0.678$. Incidentally, the maxima of all the curves fall on a straight line drawn from the right centre ($A = 0$; $\sqrt{g} = 0$) to the points $A = \pm \sqrt{(k^2 - 1)/3}$ and $\sqrt{g} = 1$, which are the corners of the frame of the curves for assays with dosage intervals of 1:2, and $k = 2$, but not otherwise.

Alignment of the maxima is one of several devices that can facilitate the construction, and make the task less formidable than it may superficially appear. For example, the points for particular curves need not be calculated in sets as outlined above; it is best to begin with a tabulation of L and then to calculate $\sqrt{g}$ for all points on the line $A = 1$, then for $A = 0.5$ and so on. Nor need "all points" be taken literally: interpolation can be used to lighten the arithmetic. (It is nevertheless advisable to construct a few "tracer" curves on a good number of points; this helps accuracy of flow of the other curves.) Another helpful circumstance lies in the fact that the pattern of the curves of upper limits is a mirror image of that of the curves of the lower limits, and many individual pairs of curves are mirrored; for example 50% and 200%, 82% and 122%.

VALIDITY OF ASSAY DATA

A pre-requisite for the use of the charts, and indeed for any method of estimating the mean potency and its precision, is that the disposition of the responses at the $2k$ points shall be consistent with the basic

order of magnitude of λ likely to be met with, together with considerations of the lay-out of the chart as a whole, indicate that a λ scale from 0 to 0.5, coupled with a $\sqrt{g}$ scale of 1 to 0, is suitable. It should be noted that $g = \sqrt{g} = 0$ corresponds to $C = 1$ and absolute precision, that $g = \sqrt{g} = 1$ corresponds to $C = \infty$ and only one real fiducial limit, and that $g > 1$ results in a negative value for C and absolute imprecision. A perpendicular drawn between these two scales will cut λ and $\sqrt{g}$ where the latter is numerically equal to $1 - 2\lambda$. Now as

$$\sqrt{g}/\lambda = t/I\sqrt{n(k^2 - 1)/12},$$

the point on the diagonal cut by the perpendicular must correspond to a unique value of n for a given I and k. This provides the key to the calibration of the diagonal. Replacing $\sqrt{g}$ by $1 - 2\lambda$ in the above expression, and re-arranging, we find

$$\lambda = I\sqrt{n(k^2 - 1)/12}/[t + 2I\sqrt{n(k^2 - 1)/12}],$$

from which any required n can be located on the diagonal as the intercept of the perpendicular from the corresponding λ. To locate, for instance, n's in the "no block" scale for $I = \log 2$ and $k = 2$ at the 5% probability level we substitute in

$$\lambda = 0.30103\sqrt{n}/(2t + 0.60206\sqrt{n}),$$

the t-table being entered at $n - 4$. Thus for $n = 20$ the value of λ is 0.1942; so we locate this n as the intercept from the perpendicular at this λ. In practice it is necessary to locate only a few n's by calculation and construction, the others being found by graphical interpolation.

The $\sqrt{g}$ scale must now be considered as one of the ordinates for the error curves to be drawn alongside. As the $\sqrt{g}$ scale thus acts merely as a stepping stone from the nomogram to the curves, in the finished diagram it need be neither labelled nor figured. The other ordinate shown on the finished charts is R, which is "found potency of test material as a multiple of the hypothesized potency", i.e. antilog IA, and the range chosen is 0.5 to 2, which will cover the great majority of assays. (In practice assays yielding values outside this range are generally repeated with an amended hypothesized potency.) The working scale however is the more amenable A, distributed linearly between ± 1 if $I = \log 2$, and between ± 1.357 if $I = \log 5/3$.

We require to plot A against $\sqrt{g}$ for given values of

$$A(C - 1) \pm \sqrt{(C - 1)[CA^2 + (k^2 - 1)/3]}$$

— or, more particularly, for given percentage limits of error at whatever probability level has been built into that expression and which

here happens to be $P = 0.05$. Let L be a root of the expression. Recalling that $C = 1/(1 - g)$ we can easily derive that

$$\sqrt{g} = \pm L/\sqrt{(L + A)^2 + (k^2 - 1)/3}.$$

Let us consider the curve of all combinations of A and $\sqrt{g}$ that yield an upper limit of 160%, i.e. 1.6 times the mean assay estimate, for 4-point designs ($k = 2$) and $I = \log 2$. The logarithm of 1.6, which is 0.2041, is IL. Therefore $L = 0.6780$. We can now calculate $\sqrt{g}$ for any value of A. For example:

A	$\sqrt{g}$
1.0	0.3471
0.5	0.4388
0	0.5612
−0.5	0.6675
−1.0	0.6454

The whole curve can be drawn from these five points. It is useful however to include a sixth, viz, the maximum, which always occurs at $L = -A$, and which for this curve is at $A = -0.678$ and $\sqrt{g} = 0.678$. Incidentally, the maxima of all the curves fall on a straight line drawn from the right centre ($A = 0$; $\sqrt{g} = 0$) to the points $A = \pm\sqrt{(k^2 - 1)/3}$ and $\sqrt{g} = 1$, which are the corners of the frame of the curves for assays with dosage intervals of 1:2, and $k = 2$, but not otherwise.

Alignment of the maxima is one of several devices that can facilitate the construction, and make the task less formidable than it may superficially appear. For example, the points for particular curves need not be calculated in sets as outlined above; it is best to begin with a tabulation of L and then to calculate $\sqrt{g}$ for all points on the line $A = 1$, then for $A = 0.5$ and so on. Nor need "all points" be taken literally: interpolation can be used to lighten the arithmetic. (It is nevertheless advisable to construct a few "tracer" curves on a good number of points; this helps accuracy of flow of the other curves.) Another helpful circumstance lies in the fact that the pattern of the curves of upper limits is a mirror image of that of the curves of the lower limits, and many individual pairs of curves are mirrored; for example 50% and 200%, 82% and 122%.

VALIDITY OF ASSAY DATA

A pre-requisite for the use of the charts, and indeed for any method of estimating the mean potency and its precision, is that the disposition of the responses at the $2k$ points shall be consistent with the basic

assumptions of the assay [13]. Validity tests must therefore be applied. (Strictly, they test for non-validity; validity is unprovable, but is assumed in the absence of evidence to the contrary.) For the 4-point assay there is only one test available: that for dissimilarity of the slopes of the response curves of the standard preparation (S) and test material (U). If we call the four response means S_1, S_2, U_1 and U_2, the fraction

$$\frac{S_1 - S_2 + U_1 - U_2}{2s/\sqrt{n}}$$

must, if the assay is unequivocally acceptable, be not more than the $P = 0.05$ value of t in Fisher's table, suitably entered. If it is more, the assay may or may not be acceptable; the choice turns on whether or not there is extraneous information on the characteristics of the response curves; and to expand this statement we must first consider the validity tests for the 6-point design.

Three distinct tests can be applied to a 6-point assay. Firstly, as there are three points per curve, the assumption of linearity of the combined S and U curves can be tested: non-linearity should entail a significant difference between the upper and lower "steps". Secondly, the assumed parallelism of the two curves can be tested: non-parallelism should entail a significant difference between the distances apart at the ends, i.e. between $U_1 - S_1$ and $U_3 - S_3$. Thirdly, the shapes, assumed to be identical, of the two curves can be compared: contrasting curvature should show up as a significant difference between, on the one hand, the distance apart at the middle and, on the other hand, the mean of the distances apart at the ends, i.e. between $U_2 - S_2$ and $(U_1 + U_3)/2 - (S_1 + S_3)/2$. So the three fractions that we can apply the t test to (with $n - 6$ degrees of freedom) are:

	numerator	*denominator*
for linearity:	$(U_1 - 2U_2 + U_3) + (S_1 - 2S_2 + S_3)$	$s\sqrt{72/n}$;
for parallelism:	$(U_1 - U_3) - (S_1 - S_3)$	$s\sqrt{24/n}$;
for common curvature:	$(U_1 - 2U_2 + U_3) - (S_1 - 2S_2 + S_3)$	$s\sqrt{72/n}$.

Illustration of possible dispositions of the 6 points is best done with (idealized) curves on an absolute, rather than on the customary relative,

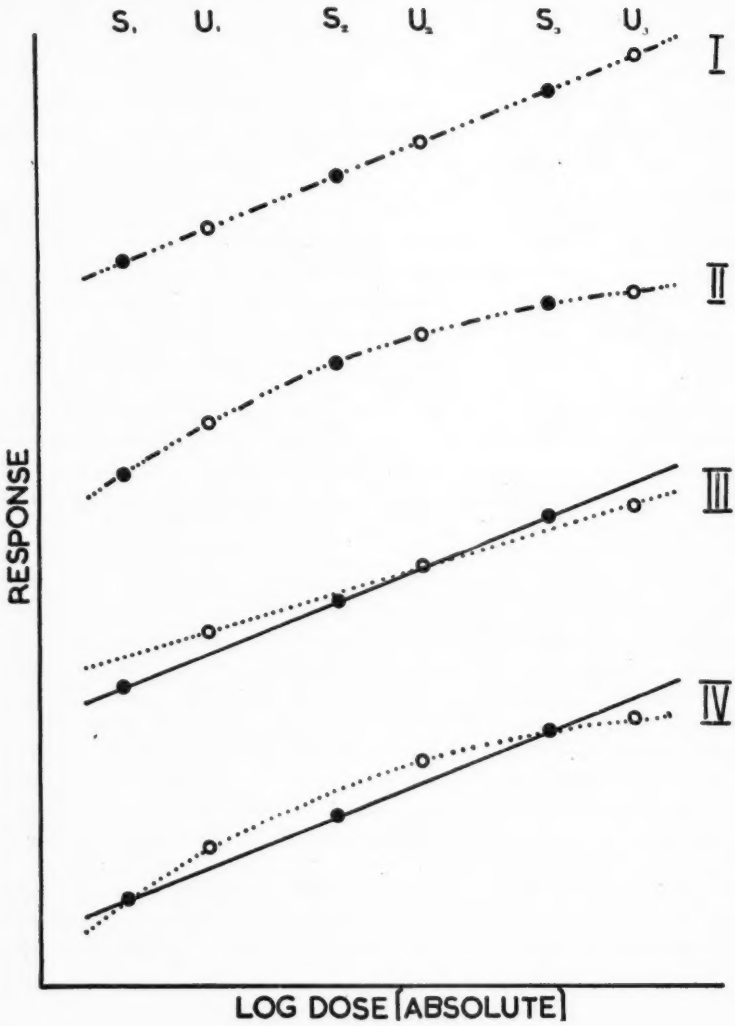


FIGURE 2.

Showing the ideal disposition of the two response curves S and U, and three possible departures from the ideal.

I coincident and linear,
II coincident and quadratic,
III non-coincident and linear,
IV non-coincident with dissimilar curvature.

log-dose scale, as in Figure 2; and in consequence "parallelism" must be transposed to "coincidence". In perfection S and U have a common rectilinear response/log-dose curve (I); all validity tests are then passed. But S and U may share a common quadratic curve (II); it is obvious that there is no fundamental objection to this situation; a finding of non-linearity is therefore not a true test of validity, but it may mean that a transformation of the metameter will improve the assay. The third possibility is that S and U have distinctive rectilinear curves (III); then "non-parallelism" emerges from the appropriate validity test; this means that the standard preparation and the test material are acting in biochemically different ways, and that the basis of comparison is unacceptable. Finally, one relation may be linear and the other curvilinear (IV), or both may be curvilinear in different ways; this results in failure of the "common curvature" test, and it disturbs the basis of comparison almost as strongly as non-parallelism. There are, of course, combinatorial possibilities, e.g., non-parallelism in association with contrasting curvature (which association, incidentally, is depicted in set IV), and they will be reflected in corresponding linkages in the validity tests.

Now the relative positions of S_1 , U_1 , S_3 and U_3 in sets II, III and IV of Figure 2 have deliberately been made identical. Therefore, if we ignore S_2 and U_2 in these sets we are left with three 4-point assays with underlying characteristic differences that would nevertheless be indistinguishable with the data available. Herein lies the ambiguity occasioned by failure of the "validity" test of a 4-point assay. Failure *may* be accounted for by the existence of a quadratic, instead of a linear, response curve, and this is not an invalidating circumstance, for not only is it biometrically unexceptionable, but it so happens that disregard of the quadratic term does not falsify the estimate—the linear equation applied to the 4 or 6 points yields the same result as the "true" quadratic equation.

If we can be confident from the inferences of other experiments that the S and U curves are coincident for this type of assay, the 4-point design may be preferred on the score of simplicity or, in certain circumstances, of efficiency (see Appendix).

OPERATION OF CHARTS

Given the response data the assayist with an appropriate chart available has six steps to take to the full assay:

(i) Calculation of mean group responses
(ii) Estimation of standard deviation
(iii) Testing for assay validity

(iv) Calculation of λ

(v) Calculation of "found/expected potency of test material" and hence of the best estimate of absolute potency in whatever metric is customary

(vi) Entering the error curves through items (iv), (v) and the total number of responses, to read off the percentage fiducial limits.

The calculation of the mean group responses needs no elaboration. Let us call the means S_1 , S_2 etc and U_1 , U_2 etc for, respectively, the standard preparation and the test material, at the separate log dosage intervals I. Estimation of the standard deviation of responses within groups is also straightforward, although if there are blocks, e.g. litters, across the groups we must eliminate the inter-block variance by means of a simple analysis of variance. Let the estimate be s. The associated number of degrees of freedom will normally be $n - 4$ for 4-point assays and $n - 6$ for 6-point assays, where n is the *total* number of responses, but appropriately fewer if blocks have been taken into account, e.g. $3n/4 - 3$ for a 4-point assay.

Having decided the validity question (see previous section), we next calculate λ: for the 4-point design with dosage steps of 1:2, we have

$$\lambda = \frac{0.6021s}{|-U_1 + U_2 - S_1 + S_2|},$$

and for the 6-point design with dosage steps of 3:5 we have

$$\lambda = \frac{0.8874s}{|-U_1 + U_3 - S_1 + S_3|}.$$

The ratio R, labelled "found/expected potency of test material" on the charts, is

$$\text{antilog } 0.30103 \frac{U_1 + U_2 - S_1 - S_2}{-U_1 + U_2 - S_1 + S_2}$$

for the 4-point "1:2" design and

$$\text{antilog } 0.2958 \frac{U_1 + U_2 + U_3 - S_1 - S_2 - S_3}{-U_1 + U_3 - S_1 + S_3}$$

for the 6-point "3:5" design. Multiplication of the expected (null hypothesis) potency by the ratio give the best estimate of potency.

Finally we enter the chart at the appropriate λ, cross to the centre scale with a straight edge aligning λ and n, and from there move vertically down into the error curves until the correct horizontal is intercepted. The upper and lower fiducial limits are read off at that point.

In proceeding along the above lines we ignore, by implication, any

TABLE 1
RESPONSES OF A CHICK ASSAY OF VITAMIN D

S_1	U_1	S_2	U_2	S_3	U_3
115	124	79	121	80	100
141	163	78	163	88	89
129	134	104	160	80	85
130	160	93	123	90	73
137	112	69	136	104*	92
112	151	86	139	87	79
140	154*	109	158	72	85
158	134	113	109	93	111
148	149	130	138	95*	101
147	165	75	122	87	105
110	144*	107	116	95	112*
122	167	98	67	75*	89
142	141	85	102	72	120
132	149	122	102	86	82*
142	161	140	99	72	89
91	150	122	123	66	92
120	150	116	47	66	124
148*	144	91	125	93	80*
129	160	94	108	99	72
154	150*	93	129	110	84
167	151	100	121	81	87
121*	155	110*	93	63	107
140	153	106	121	103	90
142	139	103*	117	102*	100
124	159	119*	90	82	68
144*	141	102	153	101*	73
127*	144	95	149*	84	110
165	130	87	144*	74*	98
TOTALS 3,777	4,134	2,826	3,375	2,400	2,597

N.B. For the meaning of the * and of the trisection of the 168 data, see text.

distortion introduced when n is not evenly sub-divided among groups, e.g. when some responses of a symmetrically planned assay are not available at the end of the experimental period. Asymmetry will effect D/B (the contributory mean response difference should be appropriately weighted), and therefore the mean assay estimate, and will effect $V(D)$ and $V(B)$ to displace the fiducial limits. In practice however the effects are generally too fine to be of importance—see next section.

SOME EXAMPLES

The data from a sizable 6-point chick assay of the vitamin-*D* potency of an oil are given in Table 1. The standard doses were 1.2, 2 and 3.3̇ units per day per chick, and the corresponding doses of the test material 1.552, 2.587 and 4.312 mg. Thus the dosage steps were 3:5 and the hypothesized potency of the oil 773 units per gram. The response was tarsal-metatarsal distance, after a 5-week experimental period. As is usual in this type of test the best metameter was found to be log-response, the exact definition being: "100 x the 2-figure log of the T.M.T. distance, in mm/100, measured on unmagnified X-ray photographs of right legs". It is this metameter that is listed in Table 1. For the purpose of illustration, assays and fiducial limits have been calculated in full (expression (1) being used for the limits) and the latter also obtained on the appropriate chart for (i) all 168 responses:

TABLE 2
ESTIMATION OF ERROR FROM THE ASSAY DATA IN TABLE 1

Selection of responses	n	Mean Estimate (units/g)	λ	Percentage Fiducial Limits at $P = 0.05$.	
				Computed	By chart
All items	168	1000	0.150	89.9–112.2	89–113
First 96	96	1081	0.158	86.0–118.9	86–119
Next 48	48	871	0.139	82.9–122.9	83–123
Final 24	24	981	0.149	74.6–143.3	74–145
Random pairs .	12	1252	0.154	60.6–284.9	61–280
Random pairs .	12	954	0.168	54.6–251.7	55–250
Random pairs .	12	856	0.149	57.7–192.1	57–193
Random pairs .	12	641	0.127	57.7–153.5	57–154

(ii) the first 96; (iii) the next 48; (iv) the final 24; and (v) some randomly selected pairs of responses from each of the six groups ($n = 12$). There is, of course, a considerable number (in fact, nearly 3×10^{15}) of possible groupings of $n = 12$ from the totality of results; it was thought interesting to select, by means of tables of random numbers, four different ones (i.e. no one response in more than one selection) as examples of small assays. The results are juxtaposed in Table 2. It should be mentioned that the graphical extraction of the errors was done visually

on quite small-scale charts (by an independent observer) without any special effort to get "fine" readings.

To illustrate the degree of approximation attainable by the use of the charts for asymmetric assays, the three "sub-assays" referred to above were artificially damaged by the deletion of 7 responses from each. The deletions, picked out from tables of random numbers, are starred in Table 1. For the smallest sub-assay, as might be expected, the asymmetrizing damage is greatest. Consider the medium sub-assay, originally with 8 responses per group, and now with $6 + 7 + 6 + 8 + 7 + 7$. To evaluate we begin by calculating the new group averages, and the overall standard deviation (now with 35 degrees of freedom). We find:

$$S_1 = 142.00;\ U_1 = 150.29;\ S_2 = 100.00;\ U_2 = 107.63;\ S_3 = 87.86;$$

$$U_3 = 94.86;\ s = 17.75.$$

These data are then used to apply the validity tests described in the preceding section: it emerges that, of the three, only what we may call the pseudo validity test, that for linearity, gives an abnormal result; the probability of random deviations' from a linear response curve being so large is apparently less than 0.02. (In the whole assay of 168 responses, the corresponding probability is between 0.1 and 0.2.) We can therefore accept the assay as meaningful.

All steps to this point are essential, no matter how the error evaluation is to be tackled. But from here onward there are alternative routes. To find the exact mean estimate and its limits we evaluate B, $V(B)$, D and $V(D)$, with due attention to the weighting necessitated by the group inequalities. Then the estimate is calculated in the ordinary way, and the fiducial limits from

$$A(C - 1) \pm \sqrt{CA^2(C - 1) + Ct^2V(D)/B^2},$$

where A and C have the meanings assigned in the early part of this paper. The alternative is to evaluate B and D from the unweighted averages:

$$B = (142.00 + 150.29 - 87.86 - 94.86)/4 = 27.39$$

$$D = (150.29 - 142.00 + 107.63 - 100.00 + 94.86 - 87.86)/3 = 7.64;$$

to calculate the mean potency estimate from B and D:

$$773 \times \text{antilog } ID/B = 773 \times \text{antilog } 0.2219 \times 7.64/27.39$$

$$= 773 \times 1.153 = 891 \text{ units}/g;$$

and to make the one further calculation,

$$\lambda = sI/B = 17.75 \times 0.2219/27.39 = 0.144.$$

We then enter the appropriate chart with this λ and $n = 41$ and read off the error range as already described. ("Found/expected potency of test material" is here 1.153.) The fiducial limits at $P = 0.05$ are so found to be 81 – 126%.

A comparison of these alternative treatments for the three asymmetrized sub-assays is given in Table 3:

TABLE 3
A COMPARISON OF EXACT AND APPROXIMATED EVALUATIONS OF ASYMMETRIC ASSAYS. (SUB-ASSAYS FROM TABLE 1 WITH STARRED RESPONSES DELETED)

n	Exact		Approximated	
	Mean (u/g)	$P = 0.05$ limits	Mean (u/g)	$P = 0.05$ limits (by chart)
96–7	1089	85.6–119.7%	1082	85–120%
48–7	891	81.0–125.9%	891	81–126%
24–7	913	66.8–162.8%	921	67–164%

It will be observed that there is nowhere any appreciable loss sustained by the replacement of the exact and laborious evaluations by the approximate method of arriving at the mean estimate and the use of the chart to obtain the fiducial limits. This is not to suggest, of course, that gross asymmetry need never be taken into account; on the contrary, it clearly must be, the important inference is that the frequently encountered slight asymmetry is biometrically negligible.

APPENDIX

I: EFFICIENCY OF ASSAY DESIGNS

Just as the invariance of an estimate of any quantity can be used as a measure of relative value, so the reciprocal of the square of the fiducial range of an assay is a measure of its efficiency and has been suggested as a routine appendage to assay results. Finney [12] has discussed the influence of design on efficiency, and the following treatment of 4- and 6-point designs is an adjunct, in slightly different form, to what he has demonstrated.

Let us confine attention to assays with $A = 0$, i.e. with accurate predictions of potency. The log fiducial range is

$$\pm I\sqrt{(C-1)(k^2-1)/3};$$

so that we may take

$$\text{efficiency} \propto 3/I^2(C-1)(k^2-1).$$

This becomes, if C is expanded, and if, in the expansion, sI/B is replaced by λ,

$$\frac{n/t^2 - 12\lambda^2/I^2(k^2-1)}{4\lambda^2}.$$

It follows that the relative efficiency of the 4-point and the 6-point designs for a given n (determinate) and a given λ (a property of the experimental animals) is,

$$\text{Efficiency ratio, 4-pt: 6-pt} = \frac{n/t_2^2 - 4\lambda^2/I_2^2}{n/t_3^2 - 3\lambda^2/2I_3^2},$$

the subscripts designating the appropriate allocations. Note that, for the given n, there will be two t's because of the difference of 2 degrees of freedom associable with the residual standard deviation in the two designs $[(n-4) - (n-6)]$, a difference whose importance dwindles as n grows. The value of this efficiency expression normally depends, it is plain to see, not merely on the ratio of the two I's, but on their absolute values; in other words efficiency statements about compared designs with a 1:2 dosage step for 6 points and a 1:4 step for 4 points will not, strictly speaking, be applicable when the steps are, for instance, 3:5 and 3:8$\dot{3}$ respectively; and in general the greater the basic dosage step (I_3) the greater the efficiency ratio will be. However, for the values of n and λ normally dealt with, this influence of I is trivial. And in one special case it is wholly absent: when $I_2/I_3 = \sqrt{8/3}$, the efficiency ratio will be asymptotic to unity (the t difference being the limiting factor) whatever the values of I; moreover, in this circumstance alone the value of A is without influence on the efficiency. A further point to bear in mind is that the efficiency here defined is a function of the working probability level, which we shall take as 0.05.

The tabulation following gives some percentage efficiencies of two kinds of 4-point assay relative to a 6-point assay, with $I = \log 2$. The first kind, "halved dosage range", is a 4-point design with the same I, i.e. we can imagine the subjects from one of the end pairs of six dosage groups removed and evenly re-distributed among the remaining four

groups. The second kind, "equal dosage range", is a 4-point design with $I = \log 4$, i.e. as if the subjects from the middle pair of six dosage groups were removed and evenly re-distributed among the remaining four.

PERCENTAGE EFFICIENCY OF 4-POINT ASSAYS COMPARED WITH 6-POINT ASSAYS WHEN THE STANDARD AND TEST DOSES YIELD EQUAL RESPONSES

		Values of n:						
	λ	10	15	20	25	30	50	100
halved dosage range.	0.1	109	96	96	96	97	98	99
	0.12	98	91	93	94	95	96	98
	0.14	83	86	89	91	93	96	98
	0.16	62	78	85	88	90	94	97
	0.18	32	70	79	84	87	92	96
	0.2	—	59	73	80	84	91	96
	0.3	—	—	17	43	56	77	90
	0.5	—	—	—	—	—	16	67
equal dosage range.	0.1	138	108	104	102	102	101	100
	0.12	143	109	104	103	102	101	100
	0.14	149	111	105	103	102	101	100
	0.16	159	113	106	104	103	101	101
	0.18	172	114	107	105	104	102	101
	0.2	193	117	109	106	104	102	101
	0.3	—	146	121	114	110	105	102
	0.5	—	—	808	193	149	117	107

The lower half of the table, referring to "equal dosage range", is the more interesting: we may presume that a switch from 6 to 4 points would involve elimination of the middle points and retention of the total dosage range. The object of such a switch would be primarily to increase the amount of information per animal; and it is perhaps remarkable to observe that in fact the increase is negligible—except for very small and/or very insensitive assays. The increase is greater the poorer the "guess" at the potency, i.e. the further A is from zero, but even so it is rarely substantial, and as Finney, who has evaluated some efficiencies for $A \neq 0$, remarks, "the increase in the fiducial range . . . may be a small price to pay for data leading to adequate tests of assay validity" [12]—referring to a change *from* 4 points *to* 6 points. There may however be dosimetric or isogenic reasons in favor of the simpler design.

II: SYMBOLOGY

An explanatory reference list follows of symbols employed in this paper that are not universally accepted, although most are in some use, and several (C, M, R, S, b, g, λ) in wide use, in the statistics of bioassay.

$A = D/B$

B = mean response increment between adjacent S's and between adjacent U's

$C = 1/(1 - g)$ = Fieller's statistic. [See ref. 10]

D = mean response difference between comparable S's and U's (as U *minus* S)

I = log of the ratio of doses given to adjacent S's or adjacent U's

L = a fiducial limit of A measured from a mean of zero. Thus antilog I $(A + L)$ is a fiducial limit (ρ_L or ρ_U) of R, while antilog $(2 + IL)$ is a percentage fiducial limit of R and of the absolute potency estimate of the test material. [See ref. 13].

$M = IA = D/b$

R = antilog M = best estimate of the unknown potency ($=\rho$ = antilog $I\alpha$), in terms of the hypothesized potency of the test material = potency ratio of comparable doses of test material and standard preparation.

S = mean response of one dosage group of the standard preparation. Distinguished as $S_1 \cdots S_k$

U = mean response of one dosage group of the test material. Distinguished as $U_1 \cdots U_k$

$b = B/I$ = best estimate of the mean linear slope of the two response/log-dose curves

$g = t^2V(B)/B^2 = 1 - 1/C$ = Finney's statistic. [See ref. 12]

k = number of dosage groups in S or U (to form, therefore, a $2k$-point assay design)

n = *total* number of responses in the assay

α = parametric value of A

$\lambda = s/b$ = a coefficient of response sensitivity.

REFERENCES

1. de Beer, E. J. The calculation of biological assay results by graphic methods. The all-or-none type of response. *J. Pharmacol. Exp. Therap., 85:* 1, 1945.
2. Knudsen, L. A. and Randall, W. A. Penicillin assay and its control chart analysis. *J. Bact., 50:* 187, 1945.
3. Bliss, C. I. A revised cylinder-plate assay of penicillin. *J. Amer. Pharm. Assoc. 35:* 6, 1946.
4. Sherwood, M. B. Simple formulas for calculating percentage potency in 3- and 4-dose assay procedures. *Science, 106:* 152, 1947.

5. Goyan, F. M. and Dufrenoy, J. A graphical calculator for bioassays. *J. Amer. Pharm. Assoc., 36:* 305, 1947.
6. Koch, W. A rapid method for the evaluation of microbiological tests. *Amer. J. Clin. Pathol., 17:* 897, 1947.
7. Harte, R. A. A simple graphical solution for potency calculations of multidose assays. *Science, 107:* 401, 1948.
8. Lees, K. A. A semi-automatic calculating machine for the computation of microbiological plate assay results. *Chem. & Ind.* 378, 1949.
9. Healy, M. J. R. Routine computation of biological assays involving a quantitative response. *Biometrics, 5:* 330, 1949.
10. Fieller, E. C. The biological standardization of insulin. *J. Roy. Statist. Soc., Suppl. 7:* 1, 1940.
11. Bliss, C. I. and Cattell, McK. Biological assay. *Ann. Rev. Physiol. 5:* 479, 1943.
12. Finney, D. J. The principles of biological assay. *J. Roy. Statist. Soc. Suppl. 9:* 46, 1947.
13. Jerne, N. K. and Wood, E. C. The validity and meaning of the results of biological assays. *Biometrics, 5:* 273, 1949.

In Memoriam

Charles P. Winsor

It is our unhappy duty to report the sudden death of Dr. C. P. Winsor on April 4. Charles Winsor was born in 1895 and was educated at Harvard University, originally as an engineer. After seven years with the New England Telephone and Telegraph Company, his interests shifted to biological research and biometry, first at Johns Hopkins University and later at Harvard where he received his Ph.D. in general physiology in 1935. In 1938 he was appointed to the Department of Mathematics, Iowa State College and during the war worked in Princeton as a research associate under a national defense research contract. He joined the Department of Biostatistics in the School of Hygiene at Johns Hopkins in 1946, a position which he held at the time of his death. He was a charter member of the Biometric Society and first vice-president of its Eastern North American Region. As editor of "Human Biology" and as a biometrical consultant he exerted a wide influence upon both biometry and biology. He will be sorely missed by his many friends and colleagues.

QUERIES

GEORGE W. SNEDECOR, Editor

88 **QUERY:** Nathan Mantel writes as follows: I believe an error crept into your reply to Query 84 in the September 1950 issue of "Biometrics". In the example you consider at the bottom of page 299 you show the following probabilities, pointing out that for this sample the exact test is closer to the uncorrected chi-square test:

Chi-square, $P = 0.038$

Corrected chi-square, $P = 0.096$

Fisher's exact test, $P = 0.046$

For the two chi-square tests, the alternative hypothesis is merely that the mortalities are different; but for the form in which you made the exact test, the alternative hypothesis is that the mortality in the treated group is greater than that in the untreated. For making the desired one-sided comparison, Fisher took advantage of the fact that chi, with one degree of freedom, is normally distributed. You can make the comparison on the basis of a two-sided test by including in the exact test all of the more extreme cases; namely, 1, 2, 6 and 7 deaths. It happens in your example that the resulting probability, 0.092, is double that for 1 and 2 deaths, making the result almost the same as that got from corrected chi-square.

ANSWER: Thank you for calling attention to this blunder. It is one which I have made before, so I am glad to be set right.

89 **QUERY:** In answering Query 85 (Biometrics 6:435, December 1950), Bliss and Winsor indicated that the range could be used instead of the standard deviation. Is this always true? If so, we can save a lot of time now used in calculating standard deviations.

ANSWER: It is always true that the range in a sample can be used for estimating σ in the normal population from which the sample is drawn, but such use would ordinarily be uneconomical. This subject is discussed by Dixon and Massey in chapters 15 and 16 of their book, "Introduction to Statistical Analysis." The copy I have is an offset edition, but McGraw-Hill has just announced publication by their firm.

In Query 85, thousands of samples had been taken, each consisting of 3 readings on some automobile operator's blood. For purposes of statistical examination, the samples cost nothing. As was pointed out, in samples of 3 the efficiency of the range, as compared to the standard deviation, is 99%. This means that, for estimating σ, 100 ranges give the same information as 99 standard deviations. Bliss and Winsor observed that for the purposes described it would be a waste of time to calculate standard deviations. Similar situations may arise in quality control and in establishments where multiple determinations are made as routine procedure.

But in most researches, data are expensive as compared to the cost of calculating the standard deviation. Suppose for example that a chemical determination costs \$2 in time, materials and equipment. From tables of the efficiency of the range it is learned that in samples of 7 this efficiency is 91%. The meaning is that for estimating σ the standard deviation in samples of 7 determinations is about as informative as is the range in samples of $7/0.91 = 8$ (approximately). So, for the same amounts of information the calculation of s at a cost of 10 or 15 cents is equivalent to one additional chemical determination at \$2.

Habitually, I use the range for making rough estimates of σ. On the basis of these estimates I am often able to advise my clients that significance or non-significance is obvious. But when decisions are close and observations are expensive, the computation of s is ordinarily the economical procedure.

THE BIOMETRIC SOCIETY

Members of the Society will be interested in the "Directory of Scientific Organizations" recently published in Paris by UNESCO. This gives a useful resume of each organization arranged alphabetically and starting with The Biometric Society.

Dr. Lloyd C. Miller, Director of Revision of the U. S. Pharmacopeia, has accepted the Chairmanship of the Committee on Sustaining Members. Those on his committee who will be responsible for inviting institutions to Sustaining Membership in ENAR are the following: E. B. Carter, A. C. De Graff, M. W. Green, W. E. Hambourger and B. L. Oser.

ENAR: By unanimous vote of the Council of the American Statistical Association the Eastern North American Region became the first organization to affiliate with the ASA under its new constitution, which went into effect in 1948.

The following committees have been appointed in ENAR: Nominating—P. T. Bruyere (Chairman), R. L. Anderson, L. N. Hazel and L. C. Miller; Membership—Boyd Harshbarger, Chairman, and 16 members; Regional Advisory Board—J. Tukey, Chairman, and 30 members; and Joint Meeting with the Institute of Mathematical Statistics—A. S. Householder, Chairman, and 4 members.

ENAR met jointly with the Institute of Mathematical Statistics at Oak Ridge, Tennessee on March 15–17, 1951. In the registration of 104, 35 were members of the Society. At the morning session on Thursday contributed papers were presented by A. E. Brandt, W. J. Youden, C. Eisenhart, R. A. Bradley and M. E. Terry, and M. L. Clark and F. X. Lynch. The afternoon session considered the use of stochastic processes in public health statistics, with papers by E. Fix and J. Neyman, and by B. G. Greenberg. The Friday morning session concerned bioassay with quantal responses, the speakers being J. Berkson, J. W. Tukey and I. Bross. In the afternoon the topic was Experimental Design and the speakers were R. C. Bose, O. Kempthorne, A. M. Dutton and L. D. Calvin. At the banquet on Friday evening E. R. McCrady served as toastmaster, and A. M. Weinberg, Director of the Oak Ridge National Laboratory, was the speaker. The closing session on Saturday morning presented three papers on multivariate analysis by S. N. Roy, A. W. Kimball and D. V. Tiedeman. Abstracts of some of these papers are being published in BIOMETRICS.

WNAR: The Annual Meeting of the Western North American Region of The Biometric Society will be held jointly with the Institute of Mathematical Statistics and the Econometric Society at the Rand Corporation, Santa Monica, California, June 14 to 16, 1951.

Région Française (*Société Française de Biométrie*): En remplacement de Monsieur LaMotte, Monsieur Georges Teissier a été élu membre du Conseil Régional, à la seance du 7 Mars, 1951.

A la réunion de la Société le 9 Mai au Laboratoire de Zoologie de l'Ecole Normale Supérieure, Paris, l'Ordre du Jour a été "Les possibilités pratiques de l'Analyse séquentielle", par Mme. Cassignol et M. P. Schutzenberger.